University of Alberta

COHERENT NOISE ATTENUATION USING RADON TECHNIQUES

by

Cristina Moldoveanu-Constantinescu ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**.

in

Geophysics

Department of Physics

Edmonton, Alberta
Fall 2006

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-22325-3
Our file Notre référence
ISBN: 978-0-494-22325-3

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Developing methods to eliminate coherent noise from seismic data is continually a subject of interest in seismic exploration. This thesis focuses on application, implementation and development of various Radon techniques to remove ground roll and multiple reflections.

The definition and classification of the Radon transform (RT) are studied. Hybrid RT is implemented in time domain and applied for ground roll attenuation. The problem is solved as an inverse problem using an iterative reweighted least squares conjugate gradient algorithm (IRLS). Sparseness constraints are introduced via Cauchy regularization norms and a right preconditioning technique is preferred to increase the computational efficiency. Synthetic and real data examples are used to illustrate the performance of hybrid Radon for ground roll attenuation.

A shifted hyperbola is adopted as a basis function for the RT to improve the approximation of the reflection events at long offsets. The shifted hyperbolic RT is applied for velocity analysis and estimation.

Acknowledgements

I think this is the most difficult part to write not because I have no one to thank but precisely because I have too many to thank and I am afraid I will leave some people out. Therefore, I would like to begin by recognizing anyone I met at some point in the almost five years spent as a graduate student at University of Alberta. Short and long encounters had some sort of impact on me.

Since research in Geophysics was the main reason of my coming and staying here, I would like to thank my supervisor, Dr. Mauricio D. Sacchi, for his continuing support. His guidance, patience and sense of humor are greatly appreciated. I could say he irradiated me with Radon... transform... and for that I'm glad! I would also like to acknowledge all members of my committee for their time and advice over the years.

My old, new, and constant friends have been a great joy to have them around. They offered me technical and scientific suggestions, cheered me up, kept my mood up or just helped me pass my time. They know who they are but I will mention them anyway (in some sort of alphabetical order), so bear with me: Ahmet Okeler, Carmen Stefanescu, Christian Weidle, Cris Matsuura, Damien Meillieux, Ersan Turkoglu, Eun Ah Lee, Gabi Constantinescu, Juefu Wang, Ligia Mateiu, Marek Welz, Natalia Gomez-Perez, Pavlo Cholach, Shawn Walker, Somanath Misra, Soner Bekleric, Uli Theune, Volkan Tuncer, Wendell Pardasie, Will Peters, Wolfgang Soyer. My time at the library wouldn't have been as pleasant without Bernice Nelson, Gail Eggen and Trevor Bergen. Edmonton cultural scene (trust me, there is one!) has been my 'oasis of serenity amid chaos'.

Special thanks go to my dear family and best friends - my sister, my mother and my grandmother. They have been always close to me in good and not so good times. There is an old saying: 'You can choose your friends but you can't choose your family'. I know I wouldn't change mine for anything in the world. Multumesc pentru tot!

To
my mother

Contents

1 Introduction 1

1.1 Motivation and scope of the thesis 3

1.2 Thesis Outline 4

2 Seismic coherent noise 6

2.1 Introduction 6

2.2 Surface waves - Ground roll 7

2.2.1 Introduction 7

2.2.2 Ground roll elimination methods 10

2.3 Multiples 16

2.3.1 Introduction 16

2.3.2 Multiple suppression methods 16

2.4 Summary 19

3 Radon transform 20

3.1 Introduction 20

3.2 Forward-Inverse Radon pair 22

3.2.1 Radon pair I 22

3.2.2 Radon pair II 22

3.3 Linear Radon transform 24

3.4 Hyperbolic Radon transform 29

3.5 Parabolic Radon transform 31

3.6 Frequency domain implementation of the Radon transform 32

3.7 Time domain implementation of the Radon transform 35

3.8 Summary 37

4 Inversion and Regularization Techniques **38**

4.1 Introduction 38

4.2 The Linear Inverse formulation 38

4.3 Regularization 41

4.4 Conjugate gradient algorithm 47

4.5 Reduction to standard form and Preconditioning 50

4.6 Summary 53

5 Hybrid Radon Transform **55**

5.1 Introduction 55

5.2 Theory 56

5.3 Synthetic data examples 60

5.4 Real data examples 68

5.5 Summary 86

6 Shifted Hyperbolic Radon Transform **87**

6.1 Introduction 87

6.2 Methodology 88

6.3 Synthetic data examples 90

6.4 Summary 98

7 Discussion and conclusions **100**

7.1 Future work 102

References **103**

List of Tables

4.1 Regularization functions and their equivalent solutions. 48

6.1 Modeling parameters for the synthetic data presented in Figure 6.4. 94

List of Figures

1.1 Parallel beam geometry. Each projection is made up of a set of line integrals through the object (After Kak and Slaney, 1988). 3

2.1 Three shot gathers illustrating different kinds of noise. (a) Marine record displaying direct arrivals, guided waves, primary reflections and long-period multiples. (b) Land data that contains direct arrivals, primary reflections and ground roll. (c) Land data containing ground roll and random noise with strong amplitudes that hamper entirely the primary reflections. (Data taken from Yilmaz, 1987). 8

2.2 Diagram of the Rayleigh waves. 9

2.3 Frequency filters: (a) High-pass filter; (b) Low-pass filter; (c) Band-pass filter. The gray section represents the frequencies that are passed 11

2.4 Band-pass filtering: (a) Land shot gather from North Africa (Yilmaz, 1987, record number 10). (b) The same shot gather after a band-pass filter with corner frequencies 15, 20, 50 and 60 Hz has been applied. (c) The difference between the original data (a) and the filtered data (b) represents the noise in the data. 12

2.5 $f-k$ filtering of synthetic data: (a) Synthetic data before filtering; (b) $f-k$ spectrum of the data before filtering; (c) $f-k$ spectrum after the filter was applied; (d) Filtered data. 14

2.6 $f-k$ filtering of land data: (a) Land data before filtering; (b) $f-k$ spectrum of the data before filtering; (c) $f-k$ spectrum after the filter was applied; (d) Filtered data 15

2.7 Diagram of multiples (After Sheriff, 1991). 17

3.1 Diagram illustrating the definition of the Radon transform. 21
3.2 Diagram illustrating the influence of the hyper-parameter μ in estimating an optimum solution. A small μ will over-fit the data, while a large μ will under fit the data. 24
3.3 Synthetic data example illustrating some aspects of the linear Radon transform. (a) Synthetic data with six linear events with different slopes: horizontal, vertical, positive, and negative. (b) Data in linear Radon domain. All events are mapped except the vertical one because it does not intersect the time axis in the time-offset domain. The solution is obtained via a conjugate gradients least squares (CGLS) algorithm with 5 iterations. (c) Reconstructed data from the previous Radon panel. All events are reconstructed except the vertical. (d) Difference between the original synthetic data (a) and the reconstructed data (c). 26
3.4 Synthetic data example illustrating the influence of the ray parameter range on the linear Radon transform. (a) Synthetic data presented in previous figure. (b) Data in linear Radon domain. The ray parameter range is smaller than in previous example, therefore the events with larger slopes are not mapped in Radon domain. The solution is obtained via CGLS with 5 iterations. (c) Reconstructed data from the previous Radon panel. (d) Difference between the original synthetic data (a) and the reconstructed data (c). 27
3.5 Synthetic data example illustrating some aspects of the linear Radon transform. (a) Synthetic data with four hyperbolic events. (b) Data in linear Radon domain. The hyperbolas are transformed into ellipses. 28
3.6 Synthetic data example comparing the result of linear and hyperbolic Radon transform when applied to data containing hyperbolic events. (a) Data in linear Radon domain. Hyperbolas are mapped to ellipses. (b) Data in hyperbolic Radon domain. Hyperbolas are focused. The solutions are obtained using a CGLS algorithm with 5 iterations. 30
3.7 Flow chart for the application of linear Radon transform in the frequency domain (After Trad, 2001). 32

4.1 Example that illustrates the three types of inverse problems by means of line fitting. The solid line represents the actual model and the dotted lines represent alternative models that fit the data. The stars denote the observations. (a) The under-determined problem ($ND < NM$; $ND = 1$, $NM = 2$). (b) The even-determined problem ($ND = NM = 2$). (c) The over-determined problem ($ND > NM$; $ND = 6$, $NM = 2$). 40

4.2 Example that compares the damped least squares solution (DLS) with the solution obtained with first derivative $\mathbf{D}_1$ as a regularization norm. 44

4.3 Example that compares the least squares solution with the solution obtained via non-quadratic regularization (Huber and Cauchy norms). . . . 46

5.1 Example illustrating the result of applying different Radon transforms to synthetic data containing hyperbolic and linear events. (a) Synthetic data containing four hyperbolic events and seven linear events. (b) Previous data in the linear Radon domain. Lines are focused while the hyperbolas are transformed into ellipses. (c) Same data in the hyperbolic Radon domain. Hyperbolas are focused and lines are not. (d) Synthetic data in the hybrid Radon domain. Both kinds of events are now focused. 57

5.2 Synthetic data example comparing the damped least squares (DLS) and the high resolution solutions for hybrid Radon transform. (a) Damped least squares solution obtained after 5 iterations of a conjugate gradient least squares (CGLS) algorithm. (b) High resolution solution obtained after 2 external iterations and 5 internal iterations each time when an iterative reweighted least squares (IRLS) conjugate gradient algorithm was applied. A sparseness constraint of the model is introduced via a Cauchy regularization norm. It can be observed that the high resolution solution is more focused than the damped least squares solution. 61

5.3 Synthetic data illustrating the difference between the damped least squares solution (DLS) and the iterative reweighted least squares (IRLS) solution. (a) Hyperbolic data recovered from the DLS panel. (b) Hyperbolic data recovered from the IRLS panel. (c) The difference between the original data (Fig. 5.1(a)) and the predicted hyperbolas from (a). (d) The difference between the original data (Fig. 5.1(a)) and the predicted hyperbolas from (b). 63

5.4 Synthetic data comparing the data prediction from the damped least squares solution (DLS) and the iterative reweighted least squares (IRLS) solution. (a) Data predicted from the DLS panel. (b) Data predicted from the IRLS panel. (c) The difference between the original data (Fig. 5.1(a)) and the predicted data from (a). (d) The difference between the original data (Fig. 5.1(a)) and the predicted data from (b). 64

5.5 Synthetic data example illustrating the performance of the hybrid Radon transform on noisy data. (a) Noisy synthetic data. Signal to noise ratio is 6 (SNR=6). (b) DLS hybrid Radon panel after 5 iterations. (c) The recovered data from the DLS hybrid Radon panel. (d) The difference between the original data (a) and the predicted data (c). 65

5.6 Synthetic data example illustrating the performance of the hybrid Radon transform on noisy data. (a) Noisy synthetic data. Signal to noise ratio is 6 (SNR=6). (b) IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations (c) The recovered data from the DLS hybrid Radon panel. (d) The difference between the original data (a) and the predicted data from (c). 66

5.7 Synthetic data example illustrating the performance of the hybrid Radon transform on noisy data. (a) Noisy synthetic data. Signal to noise ratio is 6 (SNR=6). (b) IRLS hybrid Radon panel after 2 external iterations and 20 internal iterations (c) The recovered data from the DLS hybrid Radon panel. (d) The difference between the original data (a) and the predicted data from (c). 67

5.8 (a) Field data record containing reflections, ground roll, direct arrival, and random noise. (b) DLS hybrid Radon panel after 5 iterations. (c) Data recovered from the DLS panel. 69

5.9 (a) Field data record containing reflections, ground roll, direct arrival, and random noise. (b) Recovered hyperbolic events from the hyperbolic DLS panel in Figure 5.8(b). (c) Recovered linear events from the linear DLS panel in Figure 5.8(b). (d) Data without ground roll. The linear events predicted in (c) are subtracted form the original data (a). 70

5.10 (a) DLS hybrid Radon panel after 5 iterations. (b) DLS hybrid Radon panel after 30 iterations. (c) IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. (d) IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. 71

5.11 (a) DLS hybrid Radon panel after 5 iterations. (b) DLS hybrid Radon panel after 30 iterations. (c) IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. (d) IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. 72

5.12 (a) Predicted data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. (d) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. 74

5.13 (a) Predicted data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. (d) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. 75

5.14 (a) Predicted linear data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted linear data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. (d) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. 76

5.15 (a) Predicted linear data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted linear data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. (d) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. 77

5.16 Ground roll attenuated data obtained by subtracting: (a) the predicted linear data (Fig. 5.14(a)) from the original data (Fig. 5.8(a)); (b) the predicted linear data (Fig. 5.14(b)) from the original data (Fig. 5.8(a)); (c) the predicted linear data (Fig. 5.14(c)) from the original data (Fig. 5.8(a)); (d) the predicted linear data (Fig. 5.14(d)) from the original data (Fig. 5.8(a)). $\sigma_m^2 = 0.1max|m_{DLS}|$. 78

5.17 Ground roll attenuated data obtained by subtracting: (a) the predicted linear data (Fig. 5.15(a)) from the original data (Fig. 5.8(a)); (b) the predicted linear data (Fig. 5.15(b)) from the original data (Fig. 5.8(a)); (c) the predicted linear data (Fig. 5.15(c)) from the original data (Fig. 5.8(a)); (d) the predicted linear data (Fig. 5.15(d)) from the original data (Fig. 5.8(a)). $\sigma_m^2 = 0.5max|m_{DLS}|$. 79

5.18 (a) Land data from North Africa, record number 10 from Yilmaz (1997). (b) IRLS hybrid Radon panel after 2 external iterations and 30 CG iterations every time. $\sigma_m^2 = 0.1max|m_{DLS}|$. (c) Predicted linear data from the linear panel from (b). (d) Ground roll attenuated data obtained from subtracting the linear predicted data (c) from the original data (a). 82

5.19 $f-k$ filtering of land data. (a) Land data before filtering. (b) $f-k$ spectrum of the data before filtering. (c) $f-k$ spectrum after the filter was applied. (d) Filtered data. 83
5.20 $f-k$ filtering of land data. (a) The filtered noise (Fig. 5.19(a)-Fig. 5.19(d)). (b) $f-k$ spectrum of the filtered noise. 84
5.21 (a) Land data from Middle East, record number 39 from Yilmaz (1997). (b) IRLS hybrid Radon panel after 2 external iterations and 30 CG iterations every time. $\sigma_m^2 = 0.1max|m_{DLS}|$.(c) Predicted linear data from the linear panel from (b). (d) Ground roll attenuated data obtained from subtracting the linear predicted data (c) from the original data (a). 85

6.1 Three horizontal layered model (Castle, 1994). The parameters used for modeling are $V_{P1} = 1.5$ km/sec, $V_{P2} = 3$ km/sec, and $V_{P3} = 4.5$ km/sec. . 91
6.2 The synthetic data calculated from the model presented in the previous figure. Right side shows a zoom on the second reflection to emphasize the difference between the shifted hyperbola curve and hyperbola curve. . . . 91
6.3 Synthetic data from a two layered model are shown in the upper left figure. Upper right figure shows the Radon panel obtained using HRT. The lower left figure shows the Radon panel obtained using shifted HRT (S=1.5) and the lower right figure shows the Radon panel obtained using S=1.7. 92
6.4 The synthetic data calculated from the model presented in Table 6.1 . . . 94
6.5 Shifted hyperbolic Radon panels with constant shift parameter (S). (a) Hyperbolic Radon panel (equivalent to shifted hyperbolic Radon with constant $S = 1$). (b) Shifted hyperbolic Radon panel for $S = 1.3$. (c) Shifted hyperbolic Radon panel for $S = 1.4$. (d) Shifted hyperbolic Radon panel for $S = 1.7$. 95
6.6 Diagram that illustrates the principle of a time-variant shifted hyperbolic Radon transform. 96
6.7 Shifted hyperbolic Radon panels with variable shift parameter (S). Model 1 ($S(t) = 1$). (b) Model 5 (most focused $a = 0.002$). (c) Model 6. (d) Model 8. 97
6.8 Focusing measure curve for the 10 models using a different slope. 98

6.9 Focusing measure for the shifted hyperbolic Radon transform. 99

List of symbols and abbreviations

Symbol or Abbrevation	Name or description
f	temporal frequency in Hertz (Hz)
ω	angular frequency (radians/second)
k	wavenumber (cycles/meter) spatial frequency (radians/meter)
p	projection/ray parameter/Radon parameter
r	ray position
t	traveltime in seconds
h	offset (distance between source and receiver)
τ	two way traveltime at zero offset
x	horizontal coordinate
y	vertical coordinate
δ	delta function
λ	wavelength or hyper-parameter

Symbol or Abbrevation	Name or description
ϕ, ϕ'	integration paths
R	regularization function
θ	projection angle or angle of incidence
v	velocity
Δt_{NMO}	normal moveout correction
$\mathbf{L}$	forward Radon operator
$\mathbf{L}^{-1}$	inverse Radon operator
$\mathbf{m}$	seismic data in Radon domain
$\mathbf{d}$	seismic data
$\mathbf{n}$	additive noise
$\mathbf{W_m}$	model weighting matrix
1-D	one dimensional
2-D	two dimensional
3-D	three dimensional

Symbol or Abbrevation	Name or description
CAT	Computed Axial Tomography
CG	Conjugate Gradients
CGLS	Conjugate Gradients Least Squares
CMP	Common MidPoint
CSG	Common Shot Gather
CT	Computed Tomography
DA	Direct Arrivals
FT	Fourier Transform
DLS	Damped Least Squares
FFT	Fast Fourier Transform
GR	Ground Roll
GW	Guided Waves
HRT	Hyperbolic Radon Transform
IFFT	Inverse Fast Fourier Transform
IRLS	Iterative Reweighted Least-Squares

Symbol or Abbrevation	Name or description
KL	Karhunen-Loeve
LPM	Long Period Multiples
LRT	Linear Radon Transform
NMO	Normal MoveOut
P	pressure waves
PCG	Preconditioned Conjugate Gradients
PRT	Parabolic Radon Transform
R	Reflections
RT	Radon Transform
SH	horizontal shear waves
SV	vertical shear waves
SVD	Singular Value Decomposition
V_S	shear wave velocity

Chapter 1

Introduction

Noise is omnipresent in our every day life. It is considered to be unpleasant, unexpected, or undesired and it is defined as "a disturbance that obscures or reduces the clarity of a signal". In seismology, data is collected in form of seismograms that represent a collection of traces. A trace is a recording of the Earth's response to seismic energy passing from the source, through the subsurface, and back to the receiver. Separating signal from noise is a long standing problem in science, including seismic data processing. Seismic noise can be classified as cultural, random and coherent noise. Cultural noise is the unwanted energy caused by human activity. Random noise is the energy that is not coherent between immediate traces, or in other words, the phase is not consistent from trace to trace. This type of noise can be reduced during acquisition using certain arrays of geophones, or it can partly be removed by averaging during processing. Coherent noise, as the name suggests, is energy that is coherent from seismogram to seismogram, which makes it more difficult to eliminate during acquisition or stacking. The most common types of coherent noise are the surface waves (ground roll), multiple reflections, tube waves, and air waves. In this thesis, the noise of interest is the coherent noise, ground roll and multiple reflections in particular.

A classical approach to the noise attenuation problem is to transform the data to a new domain, where distinct features are more easily separated. From the techniques used for noise removal, one of the most common is frequency-wavenumber ($f - k$) filtering. The data is transformed in frequency-wavenumber domain via a 2-D Fourier transform where signal and noise can be discriminated based on their frequency content. A frequent

problem with $f-k$ filters is aliasing, in which case the results of filtering are suboptimal.

Another conventional technique successfully used in noise attenuation is the Radon transform (RT), defined as a summation along a particular path often linear, hyperbolic and parabolic.The Radon transform has properties that make it effective for multiple suppression, ground roll filtering, and data interpolation. It was first introduced by the Austrian mathematician Johann Radon in 1917 who mathematically proved that any function and/or image can be reconstructed from a set of its projections. The mathematics developed by J. Radon represents the basis for the invention of the CAT (computed axial tomography) scan (Kak and Slaney, 1988). Needless to mention that one of the first and widely used applications of Radon transform is in the field of medical imaging/tomography and CT scanning in particular. X-ray CT scans record 2-D projections of a 3-D object. The projection of an object at a given angle θ is composed of a set of line integrals represented by the total attenuation of the beam of x-rays as it travels in a straight line through the object (Fig. 1.1). The total attenuation of a ray at position $r = x\sin\theta + y\cos\theta$, on the projection at the angle θ, is given by the line integral

$$p(r,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x\sin\theta + y\cos\theta - r)dxdy \tag{1.1}$$

where $f(x,y)$ is the image or the object that could be recovered from the projection $p(r,\theta)$. The above equation is the Radon transform of the object, also called sinogram in Computed Tomography (Kak and Slaney, 1988). To reconstruct the image, an inversion algorithm called filtered back projection is used.

Other applications include pattern recognition and image reconstruction in image processing, computer vision and remote sensing. In oceanography and remote sensing, Radon transform has been used to examine the properties of oceanic Rossby waves (Challenor et al., 1999). Rossby wave velocity is estimated from altimetry data using methods based on 2-D and 3-D Radon transform. The 3-D Radon transform offers information not only about the velocity but also about the direction of propagation of Rossby waves.

In the field of geophysics, primarily seismic imaging/tomography, Radon transform is considered to be introduced by Chapman (1978) and Schultz and Claerbout (1978). While Chapman (1978) described the Radon transform as a new method for modeling

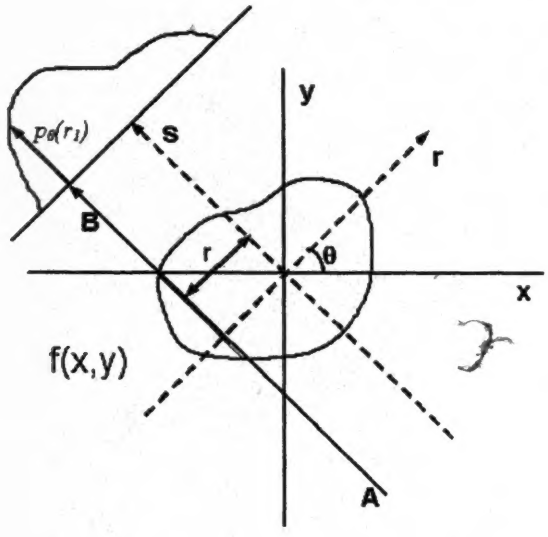

Figure 1.1: Parallel beam geometry. Each projection is made up of a set of line integrals through the object (After Kak and Slaney, 1988).

seismological data, Schultz and Claerbout (1978) used it as a tool for wave stacking and velocity estimation. As mentioned before, besides velocity analysis and velocity estimation, Radon transform has been successfully applied for multiple attenuation (Hampson, 1987; Foster and Mosher, 1992), ground roll removal (Kabir and Verschuur, 1995), and data interpolation (Hindriks and Duijndam, 1998; Trad et al., 2002; Trad, 2002b; Trad et al., 2003).

1.1 Motivation and scope of the thesis

Ground roll and multiple reflections are considered noise in seismic exploration and their elimination represents a regular subject of interest. The subject of this thesis is the application, implementation and development of various Radon techniques to remove coherent noise from seismic data. Time and frequency implementations are presented, however, this thesis focuses on noise attenuation methods in time domain associated with Radon transforms. This transformation can be implemented by means of inversion. Sparseness constraints are included in inversion via Cauchy regularization norms (Sacchi, 1997). Other regularization norms are also discussed and so the preconditioning

techniques with emphasis on the right hand preconditioning (Saad, 1996; Benzi, 2002; Calvetti, 2006). The inverse problems are solved via an iterative method called conjugate gradients least squares (CGLS) (Scales, 1987).

The Radon transform considered in this thesis as a preferred ground roll attenuation method is the hybrid Radon transform introduced by Trad et al. (2001), which is a combination of linear and hyperbolic Radon transform. The novelty consists in implementing it in time domain.

A new Radon transform shifted hyperbola equation (Castle, 1997) as a basis function is introduced. The transform is tested with synthetic data and applied for velocity analysis and estimation (Moldoveanu-Constantinescu and Sacchi, 2005). A time-variant shift parameter is implemented in the transform testing the viability of a multi-parameter Radon transform. Future applications of the shifted hyperbolic Radon transform would include multiple suppression in long-offset data and velocity estimation for long-offset data imaging.

1.2 Thesis Outline

- Chapter 1 presents an introduction of the thesis as well as the motivation and scope of the thesis. Applications of the Radon transform in other fields are briefly described.

- Chapter 2 defines coherent seismic noise showing some examples of typical noise encountered in exploration seismology. I describe several methods to remove ground roll and multiple reflections and illustrate them with examples.

- Chapter 3 focuses on time-variant and time-invariant Radon transforms in general terms and theoretical and practical aspects of the linear, hyperbolic, and parabolic transforms are enumerated. Time and frequency implementations are described and compared.

- Chapter 4 explains inversion, regularization and preconditioning techniques with emphasis on least squares inversion, Cauchy regularization and right preconditioning. The inversion is solved using an iterative reweighted least squares conjugate

gradient algorithm (IRLS).

- Chapter 5 describes the hybrid Radon transform with application to ground roll attenuation. Synthetic and real data examples are presented.
- Chapter 6 introduces the shifted hyperbolic Radon transform with application to velocity analysis and estimation.
- Chapter 7 summarizes my research and offers a discussion of the results along with possible future directions.

Chapter 2

Seismic coherent noise

2.1 Introduction

The attenuation of noise is a long standing problem in exploration seismology. In a simplistic way, noise is considered to be anything that is not wanted from the data and it could be separated into random or incoherent noise and coherent noise. Random noise, as suggested by the name, is the energy that is not coherent between immediate seismograms, or in other words, the energy does not exhibit any regularity from trace to trace. Coherent noise, on the other hand, has a distinct pattern from trace to trace. It is possible to predict how it should look on the next trace based on how it looks on the previous traces. Figure 2.1 displays three common-shot gathers (Yilmaz, 1987) from different areas that illustrate various kinds of noise. First shot gather (Fig. 2.1(a)) is a marine record from South America with an air gun source. One can distinguish the following waves: direct arrivals (DA) at about 0.25 sec, guided waves (GW) ending at 1 sec, primary reflection (R) around 0.6 sec and its long-period multiples (LPM1-LPM4) between 1 and 3 sec. Guided waves are dispersive compressional waves trapped within a water layer and traveling in the horizontal direction. They are characteristic to the shallow marine medium with hard water bottom. Long-period multiples are waves that reverberated within layers of thickness larger than a wavelength.

The second record (Fig. 2.1(b)) is a land shot gather from Alberta, the source being dynamite. Besides the reflections (R1-R5) between 1 and 3 sec and direct arrivals one can also observe ground roll energy (GR) at near offsets. Ground roll is a dispersive surface wave (Rayleigh type) traveling along the free surface, hence the name. It is characterized

by low frequency, low velocity of propagation and high amplitude.

The third shot gather (Fig. 2.1(c)) is also a land shot, but this time from North Africa, the source being a vibroseis. This is an example of a very noisy record. The reflections are practically invisible and the only things that can be noticed are the ground roll in the upper part of the record and random noise in the remaining part.

The source and the subsurface conditions differ for the three records, and also the type of the noise encountered. As we can see, there are many events that can be considered noise in exploration seismology. However, in this chapter I will focus mainly on the definition and description of Rayleigh surface waves (commonly called ground roll by the seismic community) and multiple reflections.

2.2 Surface waves - Ground roll

2.2.1 Introduction

Surface waves are waves propagating along the surface of the earth. The most typically encountered surface waves are Rayleigh and Love waves. Rayleigh waves are named after John William Strutt commonly known as Lord Rayleigh, who theoretically demonstrated the existence of this type of wave in 1885. His results and mathematical predictions were published in his paper 'On waves propagated along the plane surface of an elastic solid'.

In his paper from 1885, Lord Rayleigh investigates the behavior of waves propagated along a plane free surface of an infinite homogeneous isotropic elastic solid. He shows that the disturbance is confined to a superficial region of thickness comparable with the wavelength. Some of the conclusions he arrived at are:

- The propagation velocity of this type of surface wave is 0.9554 of the propagation velocity of the transverse waves (V_S).
- The horizontal motion disappears at a depth equal to 0.1378λ where λ represents the wavelength ($\lambda = 2\pi/f$).
- There is no finite depth at which the vertical motion vanishes.

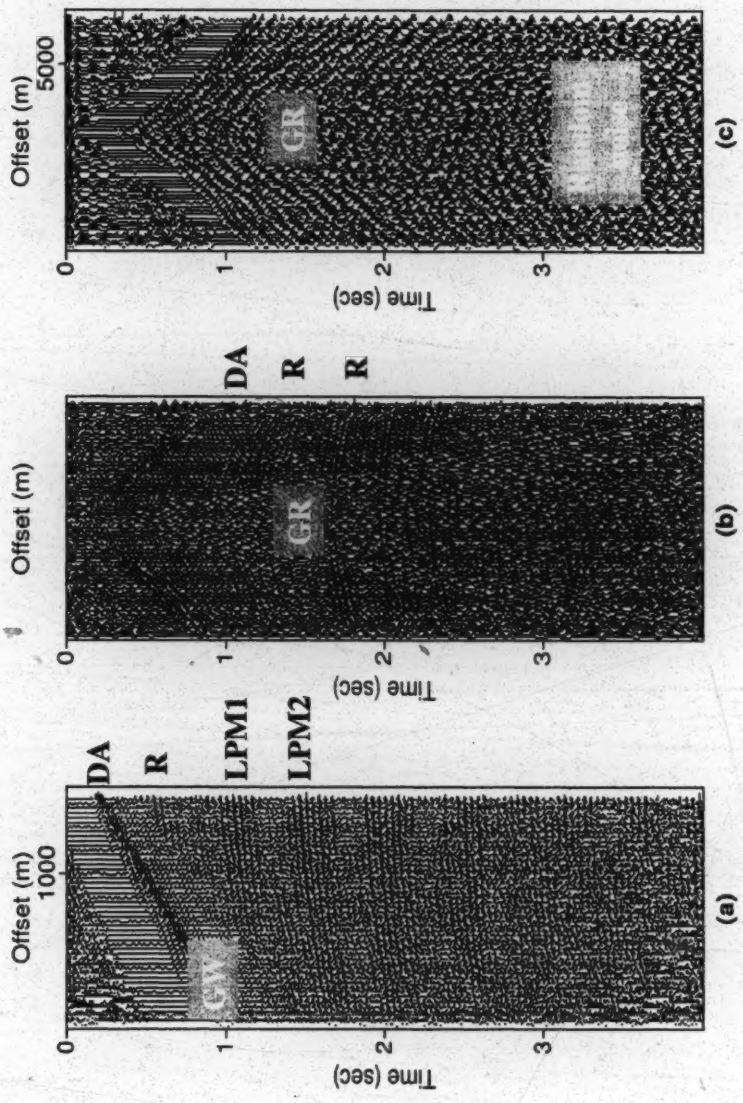

Figure 2.1: Three shot gathers illustrating different kinds of noise. (a) Marine record displaying direct arrivals, guided waves, primary reflections and long-period multiples. (b) Land data that contains direct arrivals, primary reflections and ground roll. (c) Land data containing ground roll and random noise with strong amplitudes that hamper entirely the primary reflections. (Data taken from Yilmaz, 1987).

- The motion takes place in elliptical orbits, whose vertical axis is nearly the double of the horizontal axis.

He also considered that this type of surface waves plays an important part in earthquakes, and in the collision of elastic solids, fact that was proved to be right in 1920s with the increased number of the seismographic records. To summarize, one can define Rayleigh wave as a surface wave, dispersive, with retrograde elliptical particle motion confined to the vertical plane, which includes the direction of propagation of the wave. It is actually considered that the initial retrograde elliptical motion changes with depth to no motion and then prograde elliptical motion. The propagation velocity is considered to be around $0.92V_S$, where V_S represents the propagation velocity of the transverse wave. These waves can be considered to be a combination of P and SV waves and therefore their potentials can be used to describe the Rayleigh waves. P-waves are pressure waves with a particle motion parallel to the direction of the wave propagation, while SV-waves are shear waves with a particle motion perpendicular to the direction of the wave propagation.

Figure 2.2 sketches the particle motion and direction of propagation of the ground roll. The vertical component of the motion is greater - almost double - than the horizontal component.

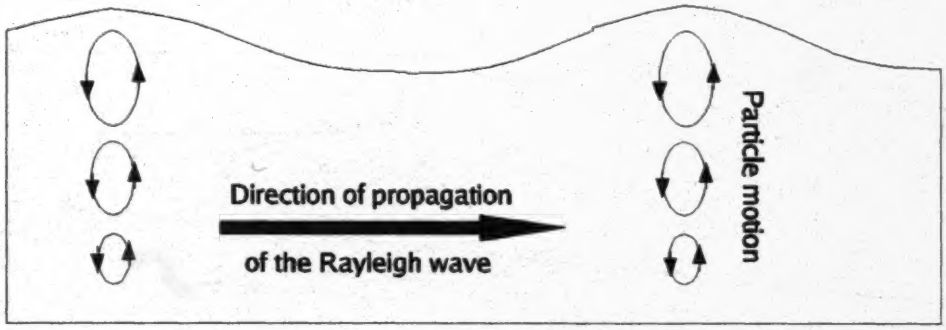

Figure 2.2: Diagram of the Rayleigh waves.

Love waves, named after A. E. H. Love, who discovered them in 1911 are another type of surface waves. They are basically surface waves of the SH type and they are not significant in exploration seismology because they usually are not generated by the energy sources used in seismic method. SH-waves are shear waves with a particle motion parallel to the ground and perpendicular to the direction of propagation or the wave.

In this thesis, the surface wave of interest is the Rayleigh wave commonly known as ground roll in exploration seismology. Unlike in earthquake seismology where they are considered useful energy, in exploration seismology they represent noise and it is desired for them to be eliminated.

2.2.2 Ground roll elimination methods

Ground roll is a dispersive type of wave that is characterized by low group velocity, low frequencies and high amplitudes. Some of these characteristics will be exploited by the different ground roll attenuation methods. One of the most popular methods used for ground roll removal is $f - k$ filtering with its offset or time variant versions. The frequency-wavenumber method is going to be explained in more detail below. Other methods for ground roll attenuation include frequency filtering, stack array method (Anstey, 1986; Morse and Hildebrandt, 1988), and filtering in the $\tau - p$ domain (Kabir and Verschuur, 1985), and wavelet domain (Deighan and Watts,1997).

Frequency filtering

As the name implies, frequency filtering attenuates the noise component from the data by removing its characteristic frequencies and passing the frequencies of the signal. Ground roll is characterized by low frequencies, while the seismic signal contains a definite range of frequencies, therefore this type of filtering would be a suitable method for ground roll suppression. This is performed in frequency domain and consists of constructing a zero-phase transfer function with a specific amplitude spectrum. The original data is firstly transformed to the Fourier domain via a Fourier transform where a frequency filter is applied. An inverse Fourier transform is then applied to the result, obtaining the filtered data. A taper is usually applied to the filter to avoid ringing, therefore the important parameters in frequency filtering are cut-off frequency and slope of the taper.

It is generally accepted that the slope at lower frequencies can be steeper than the slope at higher frequencies. There are three types of frequency filters: low-pass, high-pass, and band-pass filters. Low-pass filter also called high-cut filter, like the name suggests, rejects the high frequencies letting pass the low frequencies (Fig. 2.3(a)). High-pass (low-cut) filter would exclude low frequencies letting pass the high frequencies (gray section in Fig. (2.3(b)). However, the frequency filter typically used for ground roll attenuation is the band-pass filter (Fig. 2.3(c)) because the seismic signal is recorded in a certain range of frequencies, while seismic data usually contains low frequency noise (ground roll) and high frequency noise (ambient noise). The typical bandwidth of the signal is between 10 and 70 Hz with a dominant frequency around 30 Hz (Yilmaz, 1987).

As the figure shows this filter passes a certain range of frequency (the gray section) which represents the range of frequencies characteristic to the seismic signal.

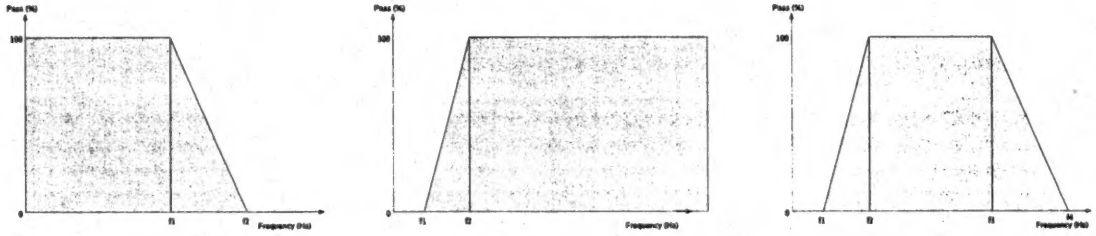

Figure 2.3: Frequency filters: (a) High-pass filter; (b) Low-pass filter; (c) Band-pass filter. The gray section represents the frequencies that are passed

In Figure 2.4 a band-pass filter with corner frequencies of 15, 20, 50 and 60 Hz was applied to land data that contains ground roll. Figure 2.4(a) illustrates shot gather number 10 from Yilmaz (1987), Figure 2.4(b) represents the band-pass filtered data, and Figure 2.4(c) illustrates the noise, in other words the difference between the original and the filtered data.

Frequency-wavenumber ($f-k$) filtering methods

These methods exploit the low frequency and low velocity characteristics of ground roll. When the data in $t-h$ domain is transformed to the $f-k$ domain the linear noise becomes isolated from the reflections. A rejection zone is defined and applied to the data elimi-

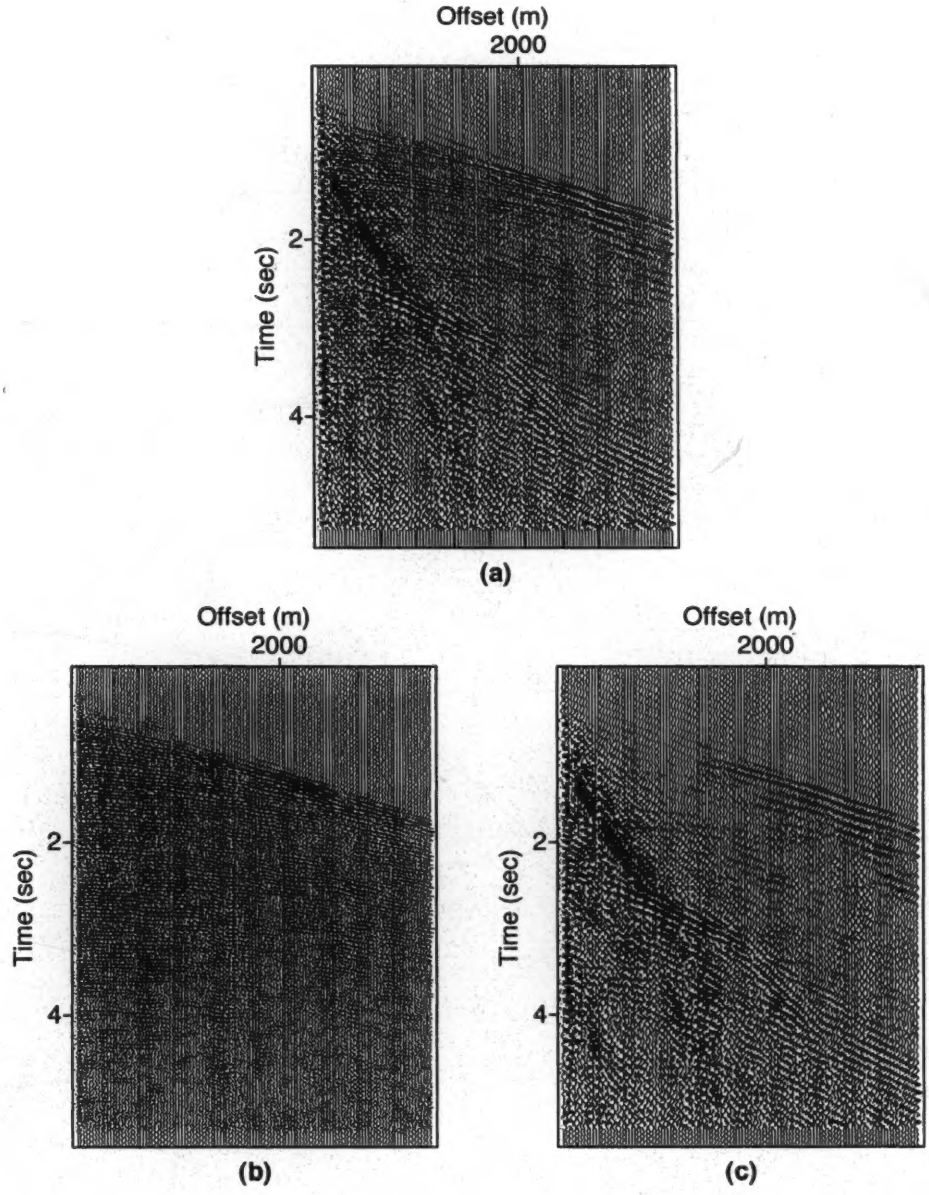

Figure 2.4: Band-pass filtering: (a) Land shot gather from North Africa (Yilmaz, 1987, record number 10). (b) The same shot gather after a band-pass filter with corner frequencies 15, 20, 50 and 60 Hz has been applied. (c) The difference between the original data (a) and the filtered data (b) represents the noise in the data.

nating the undesired energy. The degree to which the filtered record is free of ground roll depends on how well is defined the rejection zone. The shape of the rejection zone depends on the type of noise we wish to eliminate. For ground roll, a fan is commonly utilized. When defining the rejection zone, the width should be neither too narrow nor too wide and the transition between the rejection and pass zones should be smooth. Problems that frequently arise with $f-k$ filtering (Yilmaz, 1987) are spatial aliasing, temporal aliasing, and wraparound noise. To synthesize, the procedure for $f-k$ filtering is as it follows:

- Start with data in $t-h$ domain (shot gather or CMP gather).
- Apply 2-D Fourier transform (FT).
- Define a rejection zone in the $f-k$ domain and apply a mute within that zone.
- Apply 2-D inverse Fourier transform to the filtered data.

Figure 2.5 illustrates the application of an $f-k$ filter to synthetic data (Fig.2.5(a)) that contains linear noise. Figure 2.5(b) represents the $f-k$ spectrum of the input data. Wraparound energy due to spatial aliasing can be noticed in Figures 2.5(b) and 2.5(c). Spatial aliasing means insufficient sampling of the data along the space axis. The wraparound energy will not be eliminated by the dip filter, therefore there is still ground roll energy left after $f-k$ filtering (Fig. 2.5(d)).

Figure 2.6 illustrates the application of an $f-k$ filter to the field land data from the band-pass filter example.

Another method to eliminate ground roll from the seismic data is the Karhunen-Loeve (KL) transform also called principal component analysis (Liu, 1999; Sacchi, 2002; Ulrych and Sacchi, 2005; Gomez Londono et el., 2005). This transform has been successfully applied to random noise and multiple attenuation (Jones and Levy, 1987; Ulrych et al., 1988; Al-Yahya, 1991). As the name suggests, the transform is an eigenimage decomposition, similar to the singular-value decomposition (SVD).

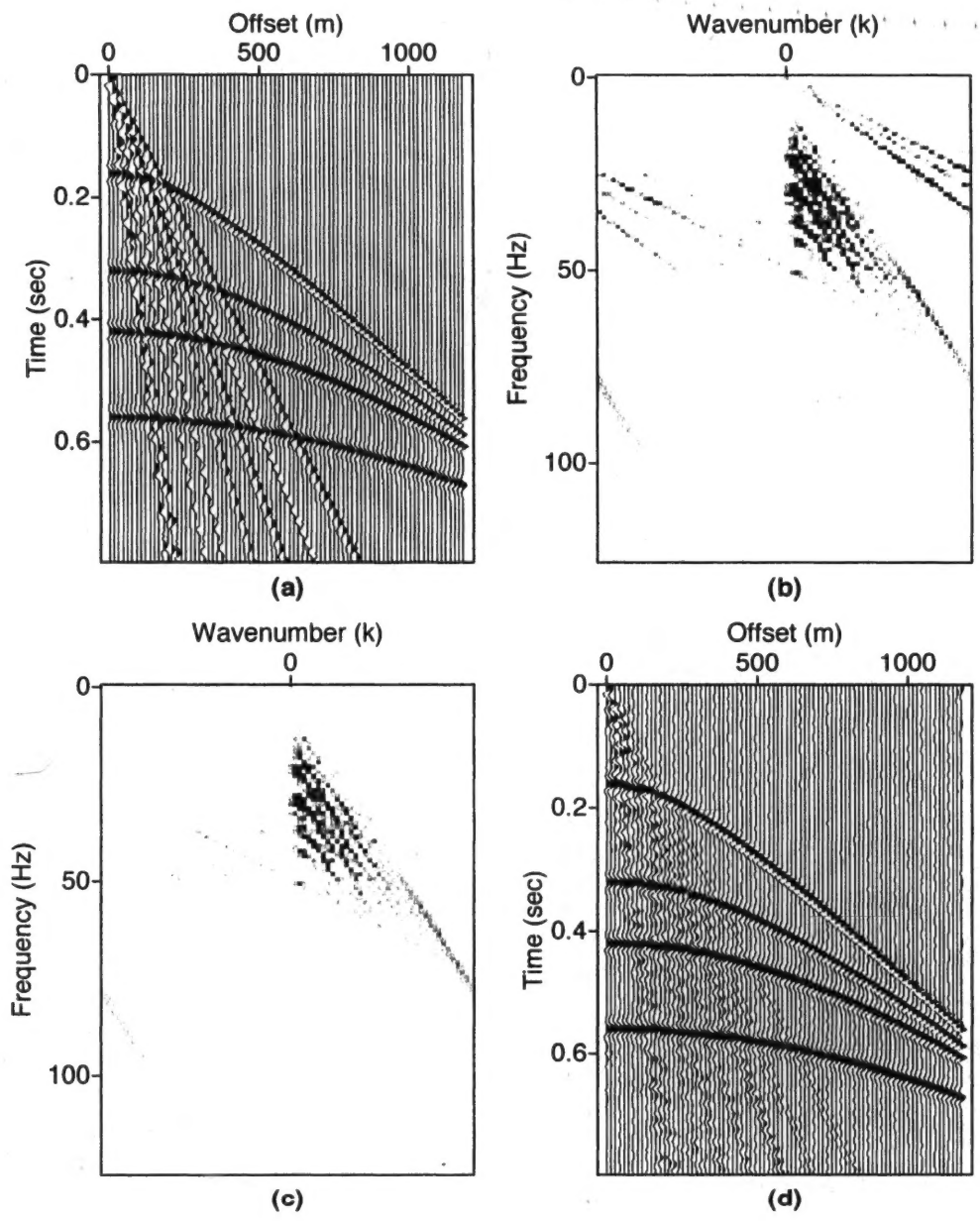

Figure 2.5: $f - k$ filtering of synthetic data: (a) Synthetic data before filtering; (b) $f - k$ spectrum of the data before filtering; (c) $f - k$ spectrum after the filter was applied; (d) Filtered data.

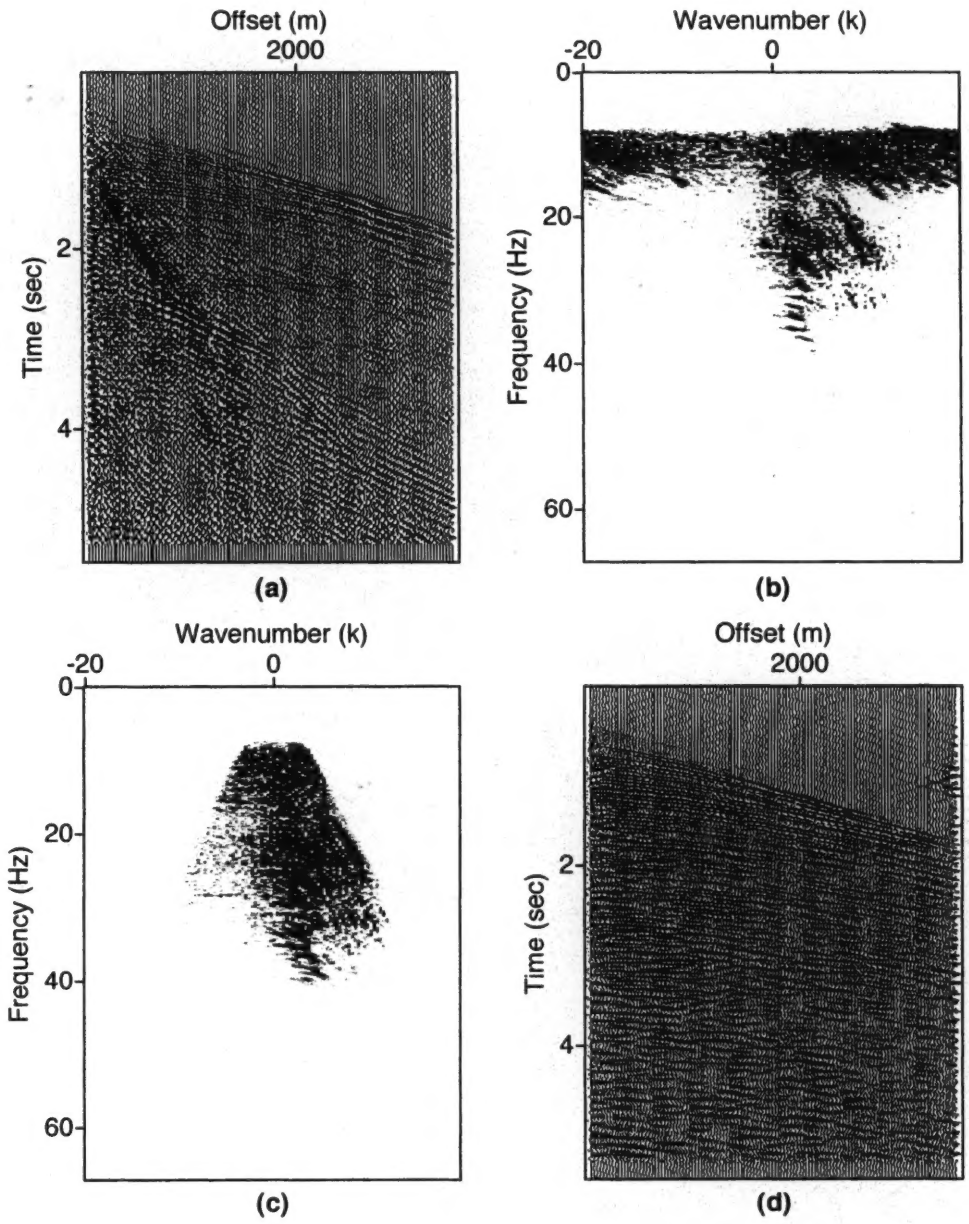

Figure 2.6: $f-k$ filtering of land data: (a) Land data before filtering; (b) $f-k$ spectrum of the data before filtering; (c) $f-k$ spectrum after the filter was applied; (d) Filtered data .

2.3 Multiples

2.3.1 Introduction

Suppression of multiple reflections is another subject of interest in seismic exploration. Multiples represent seismic energy that have been reflected back into the Earth more than once and they often interfere with the primary reflections making the interpretation of seismic data more challenging. The amplitude of the multiples depends on the reflection coefficients at each interface. Therefore, the multiple reflections visible on the seismic section and hampering the primaries are generally reflected by interfaces with a strong impedance contrast. Most common interfaces that give rise to the multiples are water bottom and the free surface. Multiples could be classified into short-path and long-path multiples (Sheriff, 1991). The name is given based on the travel path length of the multiples compared to the travel path length of the primary. Long-path multiples will travel a longer distance than the associated primaries, therefore they will also arrive at a later time making it easier to delimit between the two events. Short-path multiples usually travel a distance comparable with the primary. They often arrive soon after and they interfere with the primary reflection frequently being confused with one single event. Figure 2.4 shows different raypaths for multiples.

2.3.2 Multiple suppression methods

There are two common approaches used in seismic exploration for multiple attenuation. The filtering approach is based on exploiting certain features or physical properties that differentiate between primaries and multiples. Such features could be the periodicity of the multiples, and the moveout difference between primaries and multiples. The input data is usually transformed to a new domain where it is easier to discriminate between the different characteristics and filtering algorithms can be applied. The other approach is based on predicting the multiples using modeling and/or inversion techniques. After the multiples are predicted they are subtracted from the input data obtaining a multiple-free section.

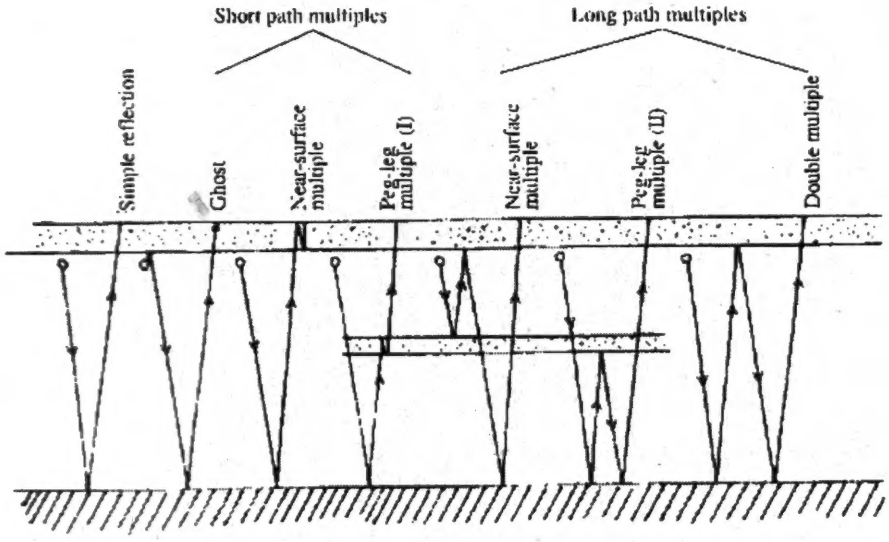

Figure 2.7: Diagram of multiples (After Sheriff, 1991).

Methods based on the periodicity of the multiples

These methods are based on the fact that multiples exhibit periodicity in contrast to the primary reflections. Robinson and Treitel (1980) use predictive deconvolution and Wiener filters to attenuate the multiples. The concept/method was introduced by Robinson (1954) and further developed by Peacock and Treitel (1969). The prediction deconvolution filters are created based on differences between the autocorrelation of input traces and desired output of the deconvolution. Two things to take into account when designing the prediction filter are the length of the filter and the prediction distance. The length of the filter is commonly considered to be equal to the length of the wavelet, and the prediction distance usually depends on the period of the multiple. Some of the assumptions made by Robinson and Treitel (1980) are stationarity and white reflectivity of the signal. At far offsets the multiples are losing their periodicity, thus the predictive deconvolution method fails. Taner (1980) applies predictive deconvolution in the radial trace space overcoming the problem of non-periodicity at far offsets.

Methods based on the moveout difference

These methods are based on the fact that a primary and a multiple reflection arriving at the same time at a geophone will exhibit a different curvature in their traveltime curves, since the two waves travel at different velocities. Common midpoint (CMP) stacking (Yilmaz, 1987), parabolic (Hampson, 1986) and hyperbolic Radon transform are examples of methods that use the moveout difference to discriminate between primaries and multiples. CMP stacking consists in applying a normal moveout (NMO) correction to the CMP gather and sum the amplitudes along the traveltime curves. The velocity used for the NMO correction is the velocity of the primary reflection. In this case, the NMO corrected CMP gather will contain flatten primaries and downwards curved multiples. If we apply a 2-D Fourier transform to the NMO corrected CMP gather, the moveout difference between the primaries and the multiples will determine them to be mapped in two different quadrants in Fourier domain (Ryu, 1982). This does not apply at near offsets, where there is no significant moveout difference.

Methods based on multiple prediction

Most of these methods make use of the wave equation to predict multiples and they are often called wave equation methods. Berryhill and Kim (1986) and Wiggins (1988) predicted multiples by propagating the wavefield down and up through the subsurface. Once multiples are predicted they are removed from the input data using an adaptive subtraction algorithm. This method works well for watter-bottom and peg-leg multiples but not for ghost multiples. The water-bottom reflectivity and topography need to be estimated. The limitation in estimating the reflectivity consists in the difficulty to discriminate between primaries and multiples that occur close together in time. Other limitations could occur due to spatial aliasing or near-offset data, and they can be overcome in the acquisition process. Other methods predict multiples based on: iterative autoconvolution in time and space (Berkhout and Verschuur, 1997; Verschuur and Berkhout, 1997); coherence measures computation (Kneib and Bardan, 1997); and inverse scattering (Weglein et al., 1992).

2.4 Summary

In this chapter, I have covered some of the theoretical and practical aspects of seismic coherent noise. Surface waves in the form of ground roll and multiple reflections have been defined. Several methods for noise attenuation have been presented. Frequency (band-pass) and frequency-wavenumber filtering are common methods for ground roll elimination. Examples illustrate these methods. Multiple suppression techniques can be classified into techniques based on certain features discrimination and techniques based on multiple prediction. Features generally used to discriminate between primary and multiple reflections are moveout difference, and periodicity of the multiples.

Chapter 3

Radon transform

3.1 Introduction

Radon transform describes line integrals in time domain as point values in Radon domain (Fig. 3.1). In other words, in geophysics Radon transform can be defined as an integration of amplitudes along a particular path. In general terms, it can be expressed as

$$\tilde{m}(\tau, p) = \int_{-\infty}^{\infty} d(t = \phi(\tau, p, h), h) dh, \tag{3.1}$$

where $\tilde{m}(\tau, p)$ represents the transformed data in Radon domain, $d(t, h)$ denotes the data in time domain or our input data, and $t = \phi(\tau, p, h)$ is the integration path. Time, offset, two-way time-intercept, and Radon parameter are represented by t, h, τ and p, respectively. The integration path $t = \phi(\tau, p, h)$ can be different and is chosen based on the problem at hand. To go back from the $\tau - p$ domain to the original $t - h$ space it is necessary to define an inverse Radon transform

$$d(t, h) = \int_{-\infty}^{\infty} m(\tau = \phi'(t, h, p), p) dp. \tag{3.2}$$

One thing to mention is that the above equation is not the exact inverse for equation (3.1), rather the two equations form a forward-adjoint pair. The forward-inverse pair can be computed via two different approaches who are classified into Radon pair I and Radon pair II (Sacchi, 1996; Li, 2001). Both approaches are going to be described in the next section. However, in the following chapters I will use only Radon pair of type II.

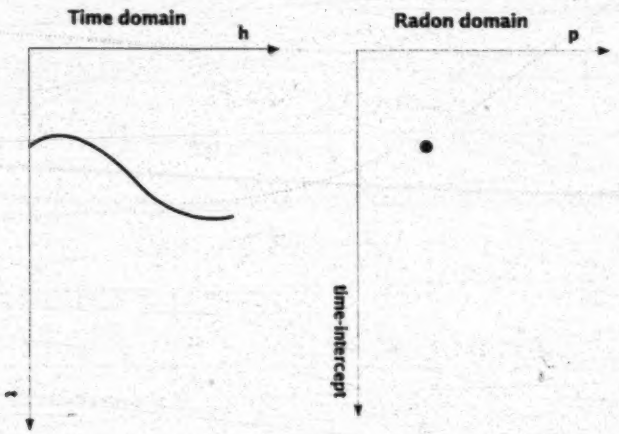

Figure 3.1: Diagram illustrating the definition of the Radon transform.

Equation (3.1) describes the forward Radon transform of a 2-D infinite and continuous function $d(t,h)$. In practice though, the data is acquired at certain values in time and space. Radon transform can therefore be expressed as a summation of amplitudes over a range of offsets $[h_{min}, h_{max}]$

$$\tilde{m}(\tau,p) = \sum_{k=1}^{Nk} d(t = \phi(\tau,p,h_k), h_k), \tag{3.3}$$

$\phi(\tau,p,h_k)$ is now a discrete function that represents the summation path and it is chosen based on the desired application. The adjoint pair then becomes

$$d(t,h) = \sum_{j=1}^{Nj} m(\tau = \phi'(t,h,p_j), p_j). \tag{3.4}$$

The usual Radon transforms used in exploration seismology integrate events of linear, parabolic or hyperbolic shape. In following sections I will describe the three classical Radon transforms with some of their applications. With the discretization of the transform come also limitations associated with finite aperture and sampling in time, space, or Radon parameter (Cary, 1998). Some of these problems can be alleviated by the techniques used to compute the discrete Radon transform. Some of these techniques will be

discussed in the following chapter.

3.2 Forward-Inverse Radon pair

To define the two kinds of forward-inverse Radon pairs we can write equations (3.3) and (3.4) in matrix form, where the model becomes a vector $\tilde{\mathbf{m}}$, data is represented by $\mathbf{d}$ and the forward and adjoint Radon operators are $\mathbf{L}$ and $\mathbf{L}^T$, respectively.

3.2.1 Radon pair I

In this case we start with the data $\mathbf{d}$ and compute the adjoint model $\tilde{\mathbf{m}}$ using the adjoint operator $\mathbf{L}^T$

$$\tilde{\mathbf{m}} = \mathbf{L}^T\mathbf{d}. \tag{3.5}$$

To go back to data domain, one can apply the operator $\mathbf{L}$ to $\tilde{\mathbf{m}}$

$$\mathbf{L}\tilde{\mathbf{m}} = \mathbf{L}\mathbf{L}^T\mathbf{d}, \tag{3.6}$$

and by inverting the operator $\mathbf{L}\mathbf{L}^T$ the data is obtained as

$$\mathbf{d} = (\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{L}\tilde{\mathbf{m}}. \tag{3.7}$$

If the inverse $(\mathbf{L}\mathbf{L}^T)^{-1}$ does not exist the pseudo-inverse $(\mathbf{L}\mathbf{L}^T)^{\dagger}$ can then replace it (Strang, 1986).

The forward-inverse Radon pair I is therefore

$$\begin{cases} \tilde{\mathbf{m}} = \mathbf{L}^T\mathbf{d} & \textbf{(Forward)} \\ \mathbf{d} = (\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{L}\tilde{\mathbf{m}} & \textbf{(Inverse)} \end{cases} \tag{3.8}$$

3.2.2 Radon pair II

In this case it is considered that data $\mathbf{d}$ is the result of the transformation $\mathbf{L}$ applied to the model $\mathbf{m}$

$$\mathbf{d} = \mathbf{Lm} \tag{3.9}$$

and can be solved as an inverse problem. If the problem is under-determined the solution is equivalent to the minimum-norm solution

$$\mathbf{m}_{mn} = \mathbf{L}^T(\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{d}, \tag{3.10}$$

and if the problem is over-determined the solution is represented by the damped least-squares solution

$$\mathbf{m}_{ls} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}, \tag{3.11}$$

where μ is a damping parameter and $\mathbf{I}$ represents the identity matrix. A classification of the inverse problems and derivation of the solutions are explained in more detail in Chapter 4 (Menke, 1984).

The forward-inverse Radon pair II with two versions, the minimum norm pair and least squares pair, is in this case

$$\begin{cases} \mathbf{d} = \mathbf{Lm} & \textbf{(Forward)} \\ \tilde{\mathbf{m}}_{mn} = \mathbf{L}(\mathbf{L}\mathbf{L}^T)^{-1}\mathbf{d} & \textbf{(Inverse)} \end{cases} \tag{3.12}$$

and

$$\begin{cases} \mathbf{d} = \mathbf{Lm} & \textbf{(Forward)} \\ \tilde{\mathbf{m}}_{ls} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d} & \textbf{(Inverse)} \end{cases} \tag{3.13}$$

As mentioned before, in this thesis I use the Radon pair II. The advantage of this approach consists in the fact that the Radon transform can be posed as an inverse problem facilitating the incorporation of model constraints. We can also consider the fitting of the noise by choosing the hyper-parameter μ. In inverse theory, the least squares solution (equation 3.13) is obtained by minimizing the cost function

$$J = ||\mathbf{Lm} - \mathbf{d}||_2^2 + \mu||\mathbf{m}||_2^2, \tag{3.14}$$

where $||\mathbf{Lm} - \mathbf{d}||_2^2$ is the misfit term and $||\mathbf{m}||_2^2$ is the model norm. These two terms and their significance will be explained in more detail in Chapter 4. If μ is too small, emphasis is put on the misfit term and therefore over-fitting the data, including the noise. If μ is too large, emphasis is put on the model norm, therefore under-fitting the data by trying to fit the desired characteristics of the model. Figure 3.2 illustrates the influence of μ in estimating the optimum solution. The best solution is determined by the χ^2 test (Tarantola, 1987; Hansen, 1998) and the trade-off parameter μ should be choose such that it gives approximately equal emphasis on both terms of the objective function.

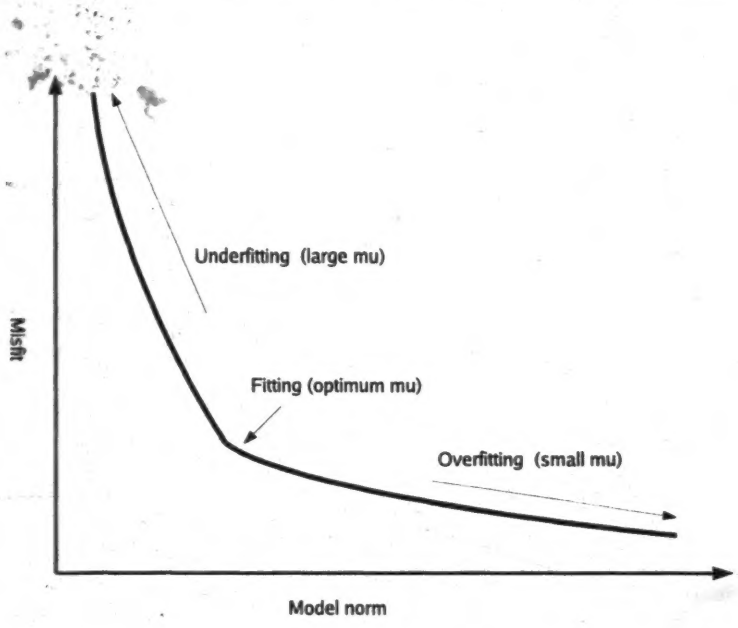

Figure 3.2: Diagram illustrating the influence of the hyper-parameter μ in estimating an optimum solution. A small μ will over-fit the data, while a large μ will under fit the data.

3.3 Linear Radon transform

Linear Radon transform (LRT), also known as *slant-stack* or $\tau - p$ transform in geophysics, was the first one introduced in seismic processing (Schultz and Claerbout, 1978). Applying a linear moveout to the data and summing amplitudes in the time-offset domain along slanted paths, a process generally called slant-stacking, is equivalent to decompos-

ing a wavefield into its plane-wave components (Treitel et al. 1982). The summation path in this case is

$$\phi(\tau, p, h) = \tau + ph, \tag{3.15}$$

where $p = \sin\theta/v$ is the ray parameter, v denotes velocity and θ represents the angle of incidence.

The forward-adjoint pair for linear Radon transform is then

$$\begin{cases} d(t,h) = \sum_{p_{max}}^{p_{min}} m(\tau = t - ph, p) & \textbf{(Forward)} \\ \\ \tilde{m}(\tau,p) = \sum_{h_{max}}^{h_{min}} d(t = \tau + ph, h) & \textbf{(Adjoint)} \end{cases} \tag{3.16}$$

Figure 3.3(a) shows a synthetic data example containing six linear events with six different slopes. Two of the events are very close, one has a negative slope, one event is horizontal ($p = 0$) and another one is vertical. Figure 3.3(b) shows the data after a linear Radon transform has been applied. It can be noticed that each of the lines in the $t - h$ domain has been transformed to peaks in the $\tau - p$ domain, except the vertical line. The vertical line does not show in the Radon domain because it does not intersect the time axis in the $t - h$ domain. The reconstructed data (Fig. 3.3(c)) from Radon domain back to the time-offset domain will not contain the vertical event either because it is not contained in the Radon domain in the first place. However, in geophysics we will not encounter events that will show as a vertical line on a record, therefore we should not be concerned about this aspect of the linear Radon transform. In a similar manner, limiting the range of the ray parameter we might limit the number of events to be mapped in the transform domain. Figure 3.4 shows the same data where the ray parameter p was chosen in a different range. In this case the two events with a ray parameter outside the used range are not imagined in the Radon domain (Fig. 3.4(b)), therefore they are not a part of the reconstructed data. It could be said that limiting the range of the Radon parameter in the $\tau - p$ transform will eliminate undesired events. However, more attention should be paid as signal and noise can be characterized by similar ray parameters.

In the new domain one can easily discriminate between different events, therefore the application of the Radon transform for data filtering is conspicuous. A transform like

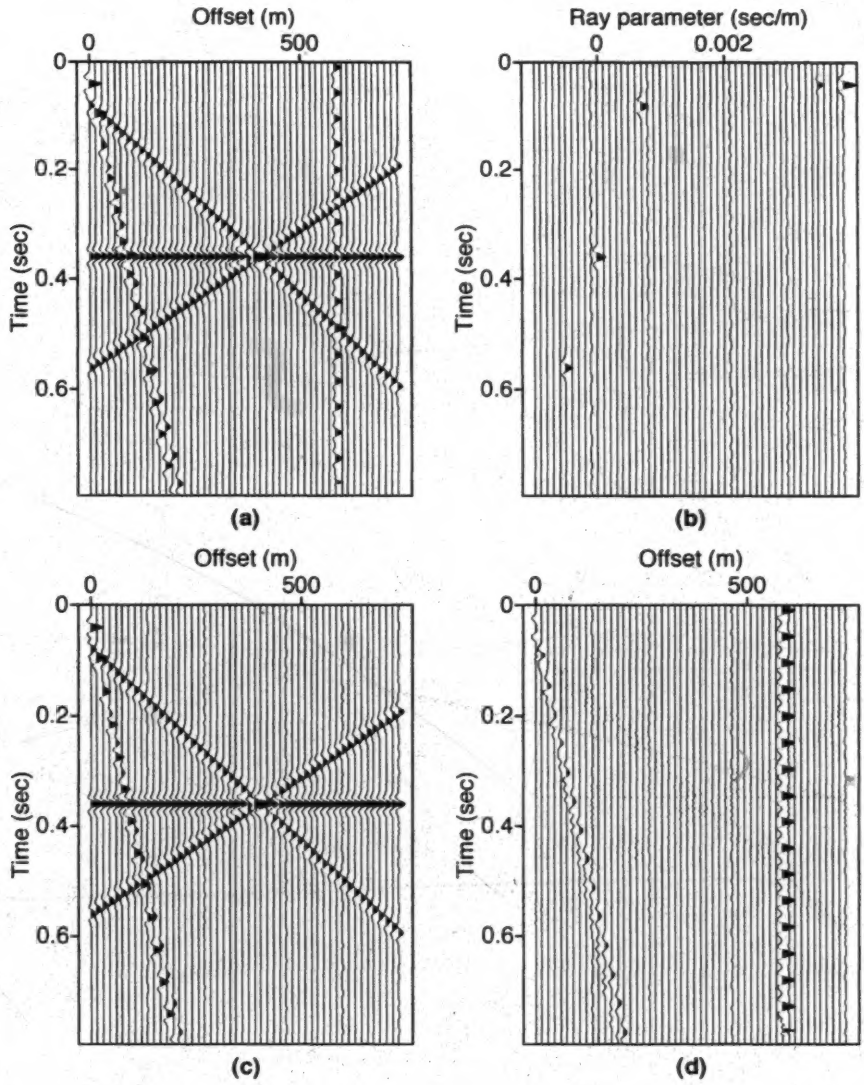

Figure 3.3: Synthetic data example illustrating some aspects of the linear Radon transform. (a) Synthetic data with six linear events with different slopes: horizontal, vertical, positive, and negative. (b) Data in linear Radon domain. All events are mapped except the vertical one because it does not intersect the time axis in the time-offset domain. The solution is obtained via a conjugate gradients least squares (CGLS) algorithm with 5 iterations. (c) Reconstructed data from the previous Radon panel. All events are reconstructed except the vertical. (d) Difference between the original synthetic data (a) and the reconstructed data (c).

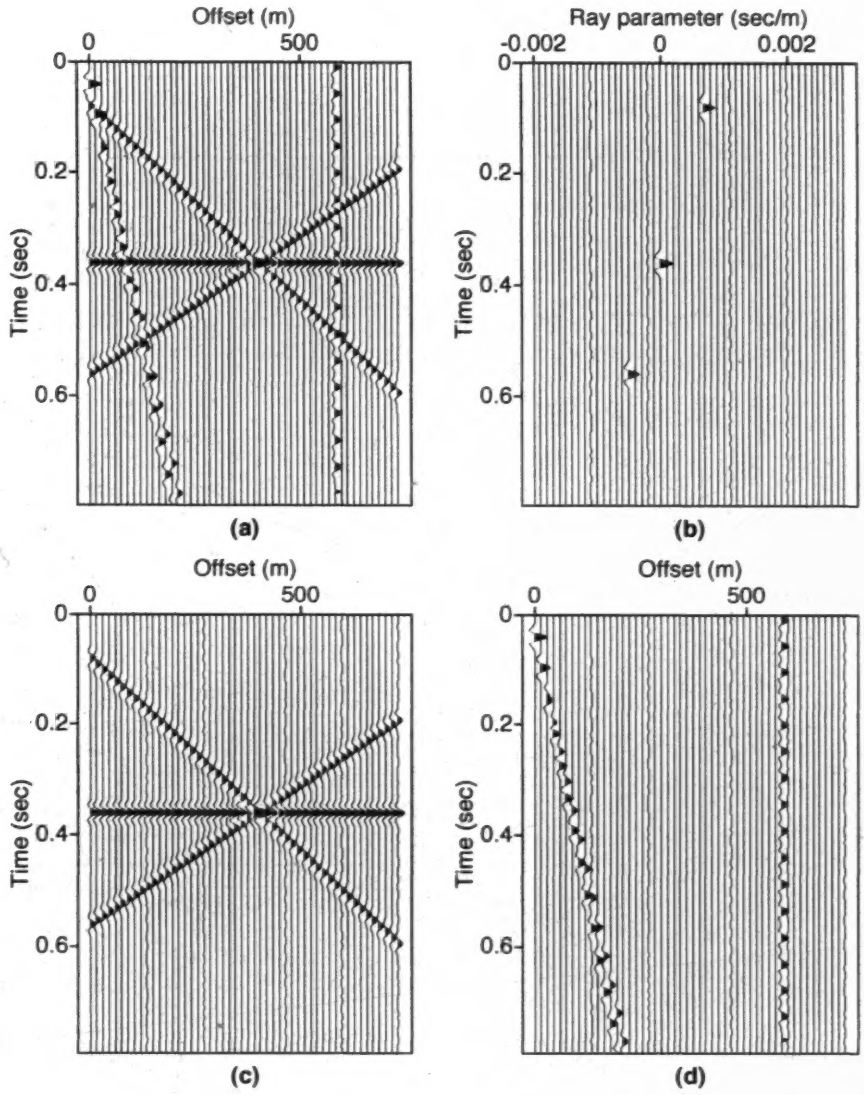

Figure 3.4: Synthetic data example illustrating the influence of the ray parameter range on the linear Radon transform. (a) Synthetic data presented in previous figure. (b) Data in linear Radon domain. The ray parameter range is smaller than in previous example, therefore the events with larger slopes are not mapped in Radon domain. The solution is obtained via CGLS with 5 iterations. (c) Reconstructed data from the previous Radon panel. (d) Difference between the original synthetic data (a) and the reconstructed data (c).

this can be used to isolate linear events in $t - h$ domain so it can be utilized to remove ground roll in common shot gathers (Yilmaz, 1987). However, reflections in common-shot and common-midpoint gathers are usually approximated by hyperbolas. A linear Radon transform will not collapse the reflections into points in the $\tau - p$ panel, but it will transform them into ellipses (Fig. 3.5).

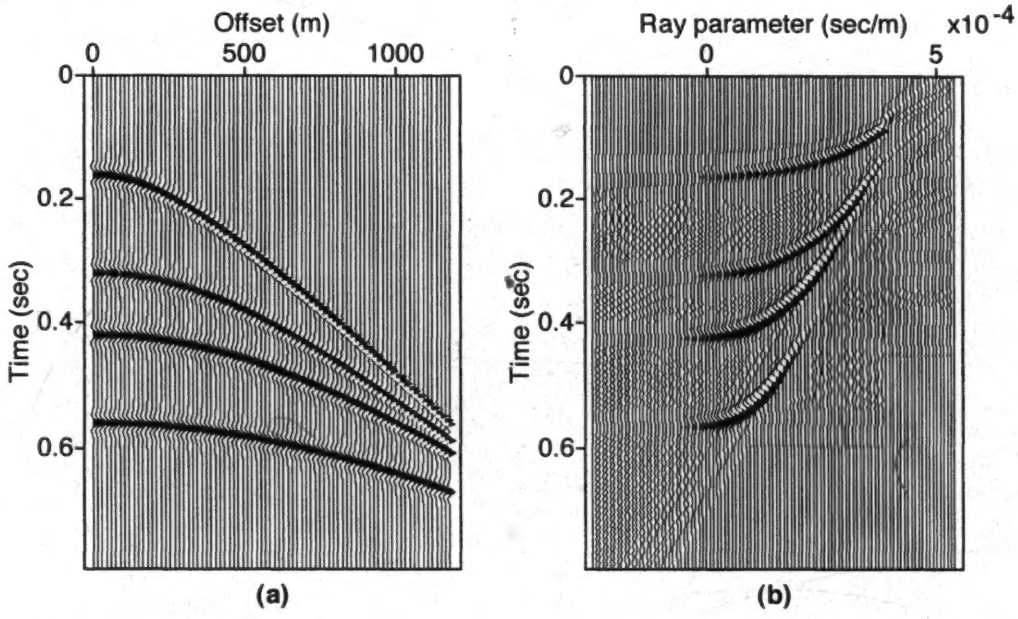

Figure 3.5: Synthetic data example illustrating some aspects of the linear Radon transform. (a) Synthetic data with four hyperbolic events. (b) Data in linear Radon domain. The hyperbolas are transformed into ellipses.

Other applications of the linear Radon transform include predictive deconvolution (Brysk et al., 1987) because multiples preserve periodicity with offset in Radon domain, data interpolation (Hindriks and Duijndam, 1998), multiple attenuation (Lokshtanov, 1993), separation of P and mode-converted SV-waves (Tatham et al., 1983; Tatham and Goolsbee, 1984; Greenhalgh et al., 1990), velocity analysis (Stoffa et al., 1982), and wavefield separation in VSP data (Moon et al., 1986; Kommedat and Tjostheim, 1989).

Linear Radon transform is time-invariant therefore possible to compute in the fre-

quency domain. An advantage of the computation in the frequency domain is that a large problem is now divided in few smaller problems, one for each frequency, that can now be solved at the same time (Fig. 3.7). This is described in more detail in section 3.6.

3.4 Hyperbolic Radon transform

As the reflections are commonly approximated by hyperbolas, the addition of a hyperbolic Radon transform was eventually imperative. Also known as *velocity-stack* this transform has been introduced by Thorson and Claerbout (1985) as a tool for velocity analysis. The new Radon parameter p is now the stacking velocity v

$$\begin{cases} d(t,h) = \sum_{v_{min}}^{v_{max}} m(\tau = \sqrt{t^2 - \frac{h^2}{v^2}}, v) \\ \\ \tilde{m}(\tau,v) = \sum_{h_{min}}^{h_{max}} d(t = \sqrt{\tau^2 + \frac{h^2}{v^2}}, h) \end{cases} , \tag{3.17}$$

all the other parameters remaining the same as in previous case. Applying this transform is equivalent to applying an NMO correction with velocity v and then summing the amplitudes over the offset. Unlike the linear Radon transform, the hyperbolic Radon transform is time-variant, therefore it cannot be computed using fast solvers algorithms (Sacchi and Porsani, 1999). Some applications of the hyperbolic Radon transform include velocity estimation and multiple attenuation.

Figure 3.6 compares two Radon panels of the synthetic shot gather illustrated in Figure 3.5(a). The first panel (Fig. 3.6(a)) represents the linear Radon panel. As mentioned before, the hyperbolas in the time-offset domain became ellipses in the $\tau - p$ domain. On the other hand, a hyperbolic Radon transform will focus the hyperbolas transforming them into points in Radon domain (Fig. 3.6(b)).

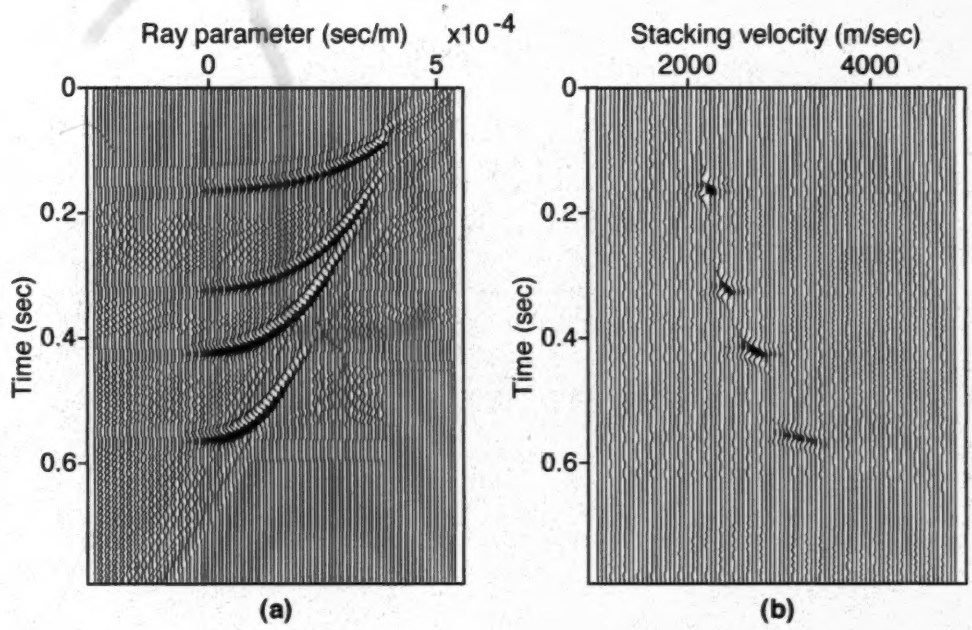

Figure 3.6: Synthetic data example comparing the result of linear and hyperbolic Radon transform when applied to data containing hyperbolic events. (a) Data in linear Radon domain. Hyperbolas are mapped to ellipses. (b) Data in hyperbolic Radon domain. Hyperbolas are focused. The solutions are obtained using a CGLS algorithm with 5 iterations.

3.5 Parabolic Radon transform

The parabolic Radon transform (PRT) was introduced by Hampson (1986). He noticed that the residual moveout of seismic reflections after a normal moveout (NMO) correction has been applied can be approximated by parabolas. The NMO correction with a velocity v_{NMO} can be written as

$$\Delta t_{NMO} = \sqrt{\tau^2 - \frac{h^2}{v_{NMO}^2}} - \tau. \tag{3.18}$$

If we apply this NMO correction to a reflection parametrized with a velocity v the travel-time then becomes

$$t = \sqrt{\tau^2 + \frac{h^2}{v^2}} - \sqrt{\tau^2 + \frac{h^2}{v_{NMO}^2}} + \tau. \tag{3.19}$$

After some calculations and rearrangements the above equation becomes

$$t = \tau + qh^2, \tag{3.20}$$

where $q = \frac{1}{2\tau}(\frac{1}{v^2} - \frac{1}{v_{NMO}^2})$ and represents the curvature of the reflections after the NMO correction.

An advantage of this transformation over the hyperbolic Radon transform is that it is time-invariant and therefore can be implemented in the frequency-offset domain. Another use of the parabolic Radon transform is proposed by Yilmaz (1989) who applies a t^2 stretching to the data. This is done by replacing $t' = t^2$ and $\tau' = \tau^2$ in the hyperbola expression that now becomes

$$t' = \tau' + \frac{h^2}{v^2}. \tag{3.21}$$

If we replace again $t = t'$ and $\tau = \tau'$ and consider $q = \frac{1}{v^2}$ the equation reduces to the parabolic form $t = \tau + qh^2$.

Fast versions of this transform were suggested by Gulunay (1990) and Kostov (1990) who observed that the parabolic Radon operator has a Toeplitz structure that enables the use of fast solvers of the Levinson recursion type. The forward-adjoint parabolic Radon pair is given by

$$\begin{cases} d(t,h) = \sum_{q_{min}}^{q_{max}} m(\tau = t - qh^2, q) \\ \tilde{m}(\tau,q) = \sum_{h_{min}}^{h_{max}} d(t = \tau + qh^2, h) \end{cases} \tag{3.22}$$

3.6 Frequency domain implementation of the Radon transform

As mentioned before, the advantage of time-invariant Radon transforms such as slant-stack or parabolic Radon transform is that they can be computed in the frequency domain thus breaking a large problem - computing the inverse of a large matrix - into several small ones (Fig. 3.7). The first step in implementing the Radon transform in the frequency-space domain is to define the discrete forward and adjoint operators. Beylkin (1987) showed that a discrete Radon transform can be computed using a least-squares technique similar to that used by Thorson and Claerbout (1985) to improve the resolution of the velocity stack. In this section I will discuss the linear Radon transform but the procedure for the parabolic Radon transform is quite similar.

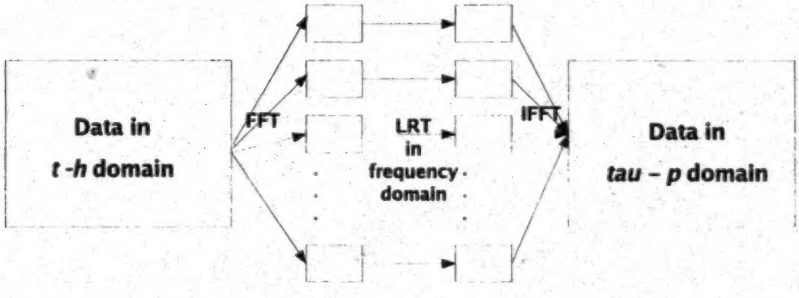

Figure 3.7: Flow chart for the application of linear Radon transform in the frequency domain (After Trad, 2001).

The discrete forward and adjoint $\tau - p$ transforms can be written as

$$\begin{cases} d(t, h_k) = \sum_{j=1}^{Nj} m(\tau = t - p_j h_k, p_j) \\ \tilde{m}(\tau, p_j) = \sum_{k=1}^{Nk} d(t = \tau + p_j h_k, h_k) \end{cases} \tag{3.23}$$

Next step is to apply 1-D Fourier transform with respect to time t to above equations. Consequently, the following expressions are obtained

$$\begin{cases} D(\omega, h_k) = \sum_{j=1}^{Nj} M(\omega, p_j) e^{-i\omega p_j h_k} \\ \tilde{M}(\omega, p_j) = \sum_{k=1}^{Nk} D(\omega, h_k) e^{i\omega p_j h_k} \end{cases} , \tag{3.24}$$

where ω is the temporal frequency. These equations can be written in matrix form as

$$\begin{cases} \mathbf{D}(\omega) = \mathbf{L}(\omega)\mathbf{M}(\omega) \\ \tilde{\mathbf{M}}(\omega) = \mathbf{L}(\omega)^H \mathbf{D}(\omega) \end{cases} , \tag{3.25}$$

where matrix $\mathbf{L}(\omega)$ is given by

$$\mathbf{L}(\omega) = \begin{pmatrix} e^{-i\omega p_1 h_1} & e^{-i\omega p_2 h_1} & \dots & e^{-i\omega p_{NJ} h_1} \\ e^{-i\omega p_1 h_2} & e^{-i\omega p_2 h_2} & \dots & e^{-i\omega p_{NJ} h_2} \\ \dots & \dots & \dots & \dots \\ e^{-i\omega p_1 h_{Nk}} & e^{-i\omega p_2 h_{Nk}} & \dots & e^{-i\omega p_{NJ} h_{Nk}} \end{pmatrix} \tag{3.26}$$

and its elements depend on the range and sampling of both the offset h and the ray parameter p. Hermitian $\mathbf{L}^H$ is the adjoint operator of $\mathbf{L}$. To compute the inverse operator we need to define an objective function J. The objective is to estimate the model $\mathbf{M}$ such that the difference between the actual data $\mathbf{D}$ and the modeled data $\tilde{\mathbf{D}} = \mathbf{L}\tilde{\mathbf{M}}$ is minimum in the least-squares sense

$$J = ||\mathbf{LM} - \mathbf{D}||_2^2 \tag{3.27}$$

To minimize the cost function J we are taking its derivative with respect to $\mathbf{m}$ and make it equal to zero

$$\mathbf{L}^H\mathbf{LM} - \mathbf{L}^H\mathbf{D} = 0 \tag{3.28}$$

We obtain the least squares solution (Lines and Treitel, 1984)

$$\mathbf{M} = (\mathbf{L}^H\mathbf{L})^{-1}\mathbf{L}^H\mathbf{D} \tag{3.29}$$

where $(\mathbf{L}^H\mathbf{L})^{-1}\mathbf{L}^H$ is the least-squares inverse of $\mathbf{L}$. The procedure is explained in more detail in the next chapter.

To summarize the implementation of the linear Radon transform in the frequency-offset domain:

- Start with data in time-offset domain $d(t, h)$.
- Apply Fourier transform along the temporal variable t, thus obtaining $\mathbf{D}(\omega, h)$.
- For each frequency ω, set up the matrix $\mathbf{L}$ which depends on the geometry of the input data gather and the ray parameter p.
- Estimate the least-squares solution for the model $\mathbf{M}(\omega, p)$.
- Apply the inverse Fourier transform to $\mathbf{M}(\omega, p)$ obtaining $m(\tau, p)$.
- Apply the desired filtering in the $\tau - p$ domain.
- Apply the inverse Radon transform obtaining the filtered data in the time-offset domain.

The implementation of the parabolic Radon transform in the frequency domain is quite similar. The discrete Radon transform in time domain becomes

$$\begin{cases} d(t, h_k) = \sum_{j=1}^{Nj} m(\tau = t - q_j h_k^2, q_j) & (\mathbf{F}) \\ \\ \tilde{m}(\tau, q_j) = \sum_{k=1}^{Nk} d(t = \tau + q_j h_k^2, h_k) & (\mathbf{A}) \end{cases}, \tag{3.30}$$

and transformed in frequency domain is

$$\begin{cases} D(\omega, h_k) = \sum_{j=1}^{Nj} M(\omega, q_j) e^{-i\omega q_j h_k} \\ \\ \tilde{M}(\omega, q_j) = \sum_{k=1}^{Nk} D(\omega, h_k) e^{i\omega q_j h_k} \end{cases}, \tag{3.31}$$

determining the matrix $\mathbf{L}$ to be

$$\mathbf{L}(\omega) = \begin{pmatrix} e^{-i\omega q_1 h_1^2} & e^{-i\omega q_2 h_1^2} & \dots & e^{-i\omega q_{NJ} h_1^2} \\ e^{-i\omega q_1 h_2^2} & e^{-i\omega q_2 h_2^2} & \dots & e^{-i\omega q_{NJ} h_2^2} \\ \dots & \dots & \dots & \dots \\ e^{-i\omega q_1 h_{Nk}^2} & e^{-i\omega q_2 h_{Nk}^2} & \dots & e^{-i\omega q_{NJ} h_{Nk}^2} \end{pmatrix} \tag{3.32}$$

The procedure can be then summarized at it follows:

- Start with data in time-offset domain $d(t, h)$.
- Apply t^2 stretching or NMO correction.
- Apply Fourier transform along the temporal variable t, thus obtaining $\mathbf{D}(\omega, h)$.
- For each frequency ω, set up the matrix $\mathbf{L}$ which depends on the geometry of the input data gather and the ray parameter p.
- Estimate the least-squares solution for the model $\mathbf{M}(\omega, p)$.
- Apply the inverse Fourier transform to $\mathbf{M}(\omega, p)$ obtaining $m(\tau, p)$.
- Undo the t^2 stretching or apply the inverse NMO correction.
- Apply the desired filtering in the $\tau - p$ domain.
- Apply the inverse Radon transform obtaining the filtered data in the time-offset domain.

3.7 Time domain implementation of the Radon transform

Notwithstanding the computational speed of the frequency domain implementation I used the time domain implementation of the Radon transform in this thesis. The advantage of the time domain implementation is that it gives more flexibility in what basis function to use for the Radon transform. Both time-variant and time-invariant RT can be performed in time domain. The first step of the implementation is also to define the discrete forward and adjoint operators (see equations 3.4 and 3.3). As previously shown, these equations can be written in matrix form as

$$\mathbf{d} = \mathbf{L}\mathbf{m} \tag{3.33}$$

and

$$\tilde{\mathbf{m}} = \mathbf{L}^T\mathbf{d} \tag{3.34}$$

where the matrix $\mathbf{L}$ represents one of the Radon operators defined in the previous sections. Additional summation paths for the Radon transform such as ellipses (Trad, 2001), apex shifted hyperbolas (Trad, 2002b), shifted hyperbolas (Moldoveanu-Constantinescu and Sacchi, 2005), allowing for a more accurate approximation of the events in time-offset domain can be also considered. The objective function is set as

$$J = ||\mathbf{Lm} - \mathbf{d}||_2^2 \tag{3.35}$$

and minimizing it with respect to $\mathbf{m}$ we obtain the least-squares solution

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{d} \tag{3.36}$$

To avoid stability problems we often add $\mu\mathbf{I}$ to the inverse operator $(\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}$ as discussed in section 3.2. *A priori* information about the model can be included in the inversion algorithm via regularization. This procedure along with different regularization functions will be described in the following chapter.

To summarize the steps of the time domain implementation:

- Start with data in time-offset domain $d(h, t)$.
- Estimate the solution $m(\tau, p)$ from the data using inversion algorithms described in Chapter 4.
- Apply the desired filtering in the $\tau - p$ domain.
- After filtering, apply inverse Radon transform obtaining the filtered data in time-offset domain.
- Subtract the filtered data from the original data if necessary.

3.8 Summary

Radon transform (RT) is a technique successfully applied in ground roll and multiple elimination as well as other applications. As in any technique, there are advantages and disadvantages. Linear Radon transform can be utilized to remove ground roll or other type of linear noise. Parabolic and hyperbolic Radon transforms are often used for multiple elimination and velocity analysis. Linear and parabolic RT are time-invariant making possible their implementation in frequency domain thus speeding up the computational process, while hyperbolic RT employs a basis function that better approximates the reflections. Some cases such as non-hyperbolic moveout, whether due to anisotropy or wide-offset acquisition geometries, still remain unresolved. High-resolution techniques introduced by Sacchi and Ulrych (1995) can be applied to overcome limited-aperture problems. These techniques will be discussed in more detail in the following chapter. Sampling and its counterpart aliasing are other problems to take into account but they are not the subject of this thesis.

In this chapter I reviewed the definition and applications of the linear, hyperbolic and parabolic Radon transform. Two different approaches to compute the inverse Radon transform are presented. Time-invariant Radon transforms are implemented in the frequency domain while time-variant transforms can be implemented in both time and frequency domain. Both implementations have been described and compared. In this thesis I use only the time domain implementation because it allows more flexibility in choosing the integration path.

Chapter 4

Inversion and Regularization Techniques

4.1 Introduction

From a geophysical point of view, inversion can be defined as a mathematical process to generate models that accurately describe the observations. Based on this definition, Radon transform can be also posed as an inverse problem and solved as such. Wherever inversion is involved we will also hear about concepts like operators, cost functions, norms, regularization, and preconditioning just to name a few. In this chapter, I will present these concepts and illustrate them with few examples, as well as some numerical algorithms for solving the inverse problem.

4.2 The Linear Inverse formulation

When talking about inversion it is imperative to also mention forward modeling. In geophysics, the forward problem

$$\mathbf{d} \simeq \mathbf{L}\mathbf{m} \tag{4.1}$$

where $\mathbf{d}$ represents the data obtained from the physical model $\mathbf{m}$ and $\mathbf{L}$ is the operator that generally contains the physics of the experiment, is often considered to be known. In our case, $\mathbf{L}$ is the operator that maps observations in transform space to data space. For instance, it can be a Fourier operator mapping data in frequency-wavenumber domain

to time-offset space, or it can be a Radon operator mapping the data from $\tau - p$ to $t - h$ domain. Data $\mathbf{d}$ is what we acquire and we frequently need to invert the operator $\mathbf{L}$ to obtain the model $\mathbf{m}$ that generated the data. The number of known data parameters (ND) and unknown model parameters (NM) will determine the type of the inverse problem (Menke, 1984). When the number of the data parameters is less than the number of model parameters ($ND < NM$) we deal with under-determined inverse problems. In this case the solution is not unique; there are practically an infinite number of solutions that fit the data. If the number of known parameters is equal to the number of unknown parameters ($ND = NM$) the problem is considered to be even-determined and the solution will be unique. Needless to mention that this situation never occurs in geophysics. The last case is the over-determined problem and it occurs when the number of the observations is larger than the number of unknowns ($ND > NM$). The solution is again non-unique.

Figure 4.1 illustrates the above classification of the inverse problems using the simple example of fitting a line $y = mx + b$, where y and x are the horizontal and vertical coordinates respectively which define the data points. The unknowns are represented by the slope m and y-intercept b, therefore the number of model parameters is equal to two ($NM = 2$). Figure 4(a) shows the case of an under-determined problem in which $ND = 1$ and $NM = 2$. As it can be seen, the number of possible lines going through one point is infinite, making the decision of which model to be chosen very difficult. A priori information about the model would be necessary in picking the right solution. The even-determined problem ($ND = NM = 2$) is depicted in Figure 4(b) and it can be noticed that there is only one line that would go through the two data points. The final example (Fig. 4(c)) illustrates the over-determined problem where the number of observations ($ND = 6$) is larger than the number of unknowns ($NM = 2$). There are several lines that would fit the data. However, model constraints could help finding the best fit line. In this case (Fig. 4.1(c)) the solution illustrated is the least squares solution. The least-squares approach considers that the data are uncorrelated and have Gaussian distribution (Menke, 1984).

However, in geophysics we often have more observations than unknowns but some of these observations are linearly dependent, which means that much of the information is superfluous making the problem a rather under-determined inverse problem. Moreover,

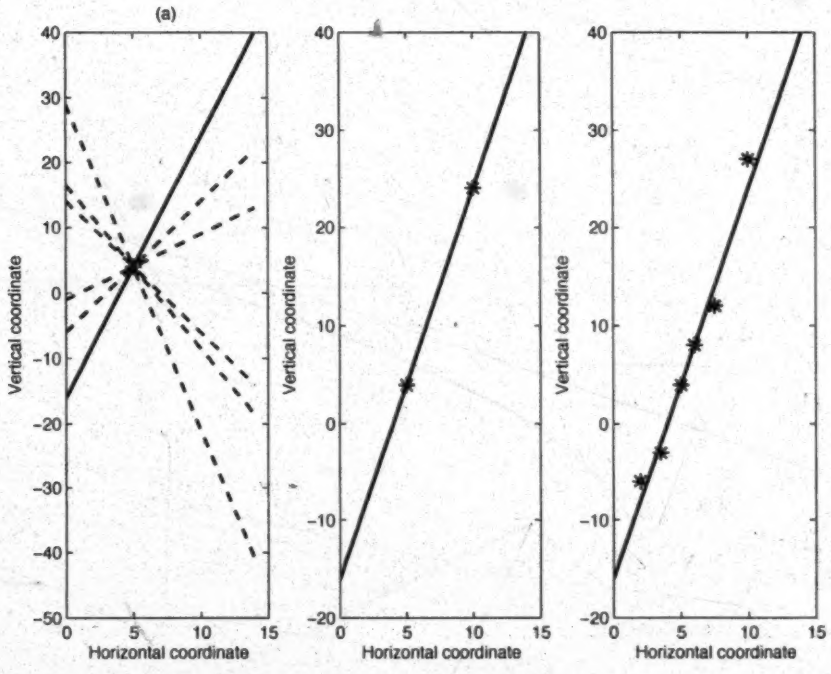

Figure 4.1: Example that illustrates the three types of inverse problems by means of line fitting. The solid line represents the actual model and the dotted lines represent alternative models that fit the data. The stars denote the observations. (a) The under-determined problem ($ND < NM$; $ND = 1$, $NM = 2$). (b) The even-determined problem ($ND = NM = 2$). (c) The over-determined problem ($ND > NM$; $ND = 6$, $NM = 2$).

the observations are discrete, not fully accurate, and sometimes noisy, in which case the equation (4.1) becomes

$$\mathbf{d} = \mathbf{Lm} + \mathbf{n} \tag{4.2}$$

where $\mathbf{n}$ represents modeling errors, additive Gaussian noise and/or missing data. The errors are assumed to be Gaussian and uncorrelated in which case the least squares method is adopted. If the errors are correlated the covariance will be included in the least-squares solution (Menke, 1984).

The geophysical inverse problems are generally considered to be ill-posed or ill-conditioned (Menke, 1984). To find the best model that fits the data we need to define a cost function, which is a mathematical expression that contains the undesired character-

istics of the unknown model. The first objective of this cost function is to minimize the difference between the model response and the data. This can be written as

$$J = ||\mathbf{Lm} - \mathbf{d}||_q^q \tag{4.3}$$

where $||.||_q^q$ represents the norm and it is also related to the desired characteristics of the solution. The most popular norm is the l_2 norm ($q = 2$) which considers Gaussian errors. This norm penalizes very large elements in the model, keeping only the small elements. If we want to consider errors that contain outliers, the l_1 norm is then more suitable because it is sensitive only to the total size of the errors, and allows the large elements to become larger and the small elements to become smaller (Huber, 1981). By minimizing the above cost function with respect to the model $\mathbf{m}$ in the least squares sense (Lines and Treitel, 1984) the following solution is obtained

$$\hat{\mathbf{m}} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{d}. \tag{4.4}$$

The above expression represents a very low resolution solution to the inverse problem. The operator $\mathbf{L}^T\mathbf{L}$ is a large matrix. If due to insufficient data or linear dependencies in the input data the matrix is close to singular (its determinant is close to zero) it will be difficult or impractical to invert. Also, although it is a predominantly diagonal matrix, sometimes the side lobes are not insignificant which may cause smearing along the Radon parameter axis. This solution is often unstable and non-unique. To overcome this, regularization is frequently applied. Regularization imposes stability on an ill-conditioned problem by including a priori information about the solution. In the next section I will define regularization and describe quadratic and non-quadratic regularization norms.

4.3 Regularization

As mentioned before, geophysical data are discrete, acquired at certain values in time and space, not always accurate, and sometimes noisy. We usually deal with ill-posed inverse problems. In this case, through inversion one can find many models that fit the data and the solution is often unstable. Imposing additional information about the model

will permit us to choose one particular solution to the inverse problem, one particular model that satisfies the data and it also stabilizes the solution. This procedure is called regularization. The cost function in this case will look like

$$J = ||\mathbf{Lm} - \mathbf{d}||_2^2 + \mu R(\mathbf{m}) \tag{4.5}$$

where μ is a hyper-parameter also called trade-off or damping parameter and defines the amount of weight given to the regularization function which is represented by $R(\mathbf{m})$. If μ is too small all or most of the weight is given to the misfit term $||\mathbf{Lm} - \mathbf{d}||_2^2$ and the computed solution will fit the data, signal and noise included. If μ is too large the weight is given to the regularization term and the determined solution will have the desired characteristics without trying to satisfy the input data. A common technique to determine the optimum damping parameter is called the L-curve or χ^2 test (Tarantola, 1987; Hansen, 1998). The regularization function is chosen depending on the problem at hand.

The simplest regularization function is $R(\mathbf{m}) = ||\mathbf{m}||_2^2$ and represents a quadratic norm of the model. The solution is then given by

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}, \tag{4.6}$$

where $\mathbf{I}$ denotes the identity matrix. Equation (4.6) is called the minimum quadratic norm solution or the damped least squares solution (DLS) (Claerbout, 1992). As it can be seen this regularization adds a constant μ to the diagonal of the operator $\mathbf{L}^T\mathbf{L}$. This is equivalent to whitening the solution (Thorson and Claerbout, 1985).

Weighting functions ($\mathbf{W}_m$) can be included to take into account prior knowledge about the model. In this case, the regularization functions becomes $R(\mathbf{m}) = ||\mathbf{W}_m\mathbf{m}||_2^2$. The solution in this case becomes

$$\hat{\mathbf{m}} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{W}_m^T\mathbf{W}_m)^{-1}\mathbf{L}^T\mathbf{d}. \tag{4.7}$$

The first order derivative operator $\mathbf{W}_m = \mathbf{D}_1$ will allow small variations between consecutive elements giving rise to a flat solution. Second order derivative operator

$\mathbf{W}_m = \mathbf{D}_2$ will allow smooth variations between sets of parameters giving rise to smooth solutions. The two derivatives are given by

$$\mathbf{D}_1 = \begin{pmatrix} 1 & -1 & 0 & \dots \\ 0 & 1 & -1 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{4.8}$$

and

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & \dots \\ 0 & 1 & -2 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{4.9}$$

Both derivative norms are quadratic and the cost functions give rise to linear systems of equations. The regularization functions do not depend on the model and if the matrix to be inverted $\mathbf{L}^T\mathbf{L} + \mu\mathbf{W}_m^T\mathbf{W}_m$ is neither large nor sparse we can adopt a direct approach to solve the problem (Tarantola, 1987). In the case of large and sparse matrices an iterative approach like steepest descent or conjugate gradient (CG) can be applied. In this thesis, the inversion is solved by means of a conjugate gradient algorithm (Hestenes and Steifel, 1952) and explained in the following section. This is a semi-iterative method (Claerbout, 1992) that is used to solve large linear system of equations (operators in this case). An interesting feature of the CG algorithm is that the operator $\mathbf{L}$ does not need to be explicitly contained in an array (matrix) and it is just a collection of computational procedures (functions or subprograms). Figure 4.2 compares the solutions found with two quadratic regularization functions.

If sparseness is expected in the solution, as in the case of Radon transform, variable regularization terms are required. Two norms that proved to be suitable for the computation of sparse models are Huber norm and Cauchy norm (Sacchi, 1997). These norms are non-quadratic and give rise to non-linear solutions.

When Huber norm (Huber, 1981) is applied the regularization function becomes

$$R(\mathbf{m}) = \sum_{i=1}^{NM} \rho(m_i), \tag{4.10}$$

where NM is the number of unknown model parameters, and

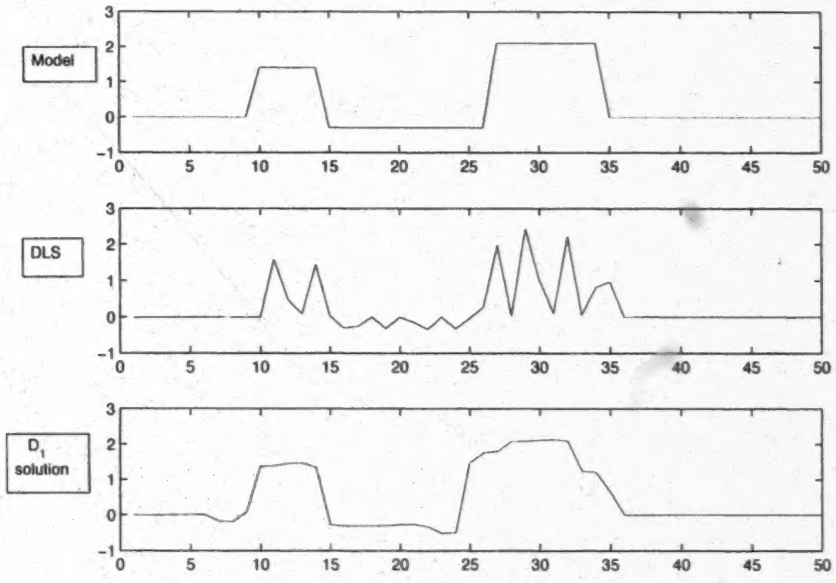

Figure 4.2: Example that compares the damped least squares solution (DLS) with the solution obtained with first derivative $\mathbf{D}_1$ as a regularization norm.

$$\rho(m_i) = \begin{cases} \frac{m_i^2}{2}, & \text{if } |m_i| \leq a \\ a|m_i| - \frac{a^2}{2}, & \text{if } |m_i| > a \end{cases}, \tag{4.11}$$

and the solution will depend on the model itself

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m}))^{-1}\mathbf{L}^T\mathbf{d} \tag{4.12}$$

where $\mathbf{Q}(\mathbf{m})$ is a diagonal matrix whose elements are

$$Q_{ii} = \begin{cases} |m_i|, & \text{if } |m_i| \leq a \\ a[\text{sign}(m_i)], & \text{if } |m_i| > a \end{cases}. \tag{4.13}$$

In the above expressions, a represents a threshold parameter that needs to be defined. The selection of a follows the rule $a = cx\sigma_m$, where c is a scalar typically considered to be between 0.1 and 1, and σ_m is a scaling factor estimated based on the assumption that the data misfit matches the power of the noise (Sacchi, 1997). When the parameter a is small compared to the unknown m the above norm behaves like an l_1 norm ($R(\mathbf{m})$).

When the Cauchy norm is applied, the regularization term is given by

$$R(\mathbf{m}) = \sum_{i=1}^{NM} \ln\left(1 + \frac{m_i^2}{\sigma_m^2}\right), \tag{4.14}$$

and again, the solution will contain the model

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m}))^{-1}\mathbf{L}^T\mathbf{d} \tag{4.15}$$

where $\mathbf{Q}(\mathbf{m})$ is again a diagonal matrix with elements equal to

$$Q_{ii} = \frac{1}{1 + \frac{m_i^2}{\sigma_m^2}} \tag{4.16}$$

and σ_m is a scaling factor of the Cauchy distribution. This scale parameter cannot be called variance because the Cauchy norm does not have a finite second order moment.

Figure 4.3 compares the least squares solution with the solution obtained using Huber and Cauchy norms as regularization functions. It can be observed that the non-quadratic regularization (Fig. 4.3(c) and (d)) gives results that resemble more the original model (Fig. 4.3(a)) as opposed to the least squares result (Fig. 4.3(b)).

We can solve the non-linear inverse problem iteratively using the Iterative Reweighted Least-Squares (IRLS) algorithm (Scales, 1987; Scales et al. 1988). The steps for the IRLS are as follows:

- Start with the damped least-squares solution (DLS)

  $\mathbf{m}_0 = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}$

- DLS is used to compute the model-dependent weighting function

- The regularization term is now used to find a new solution

  $\mathbf{m}_1 = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m}_0))^{-1}\mathbf{L}^T\mathbf{d}$

- The regularization term is updated systematic using the previous solution.

- A new solution is estimated

  $\mathbf{m}_i = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m}_{i-1}))^{-1}\mathbf{L}^T\mathbf{d}$,

  where $\mathbf{m}_i$ and $\mathbf{m}_{i-1}$ are the solutions at the ith and $(i-1)$th iteration respectively.

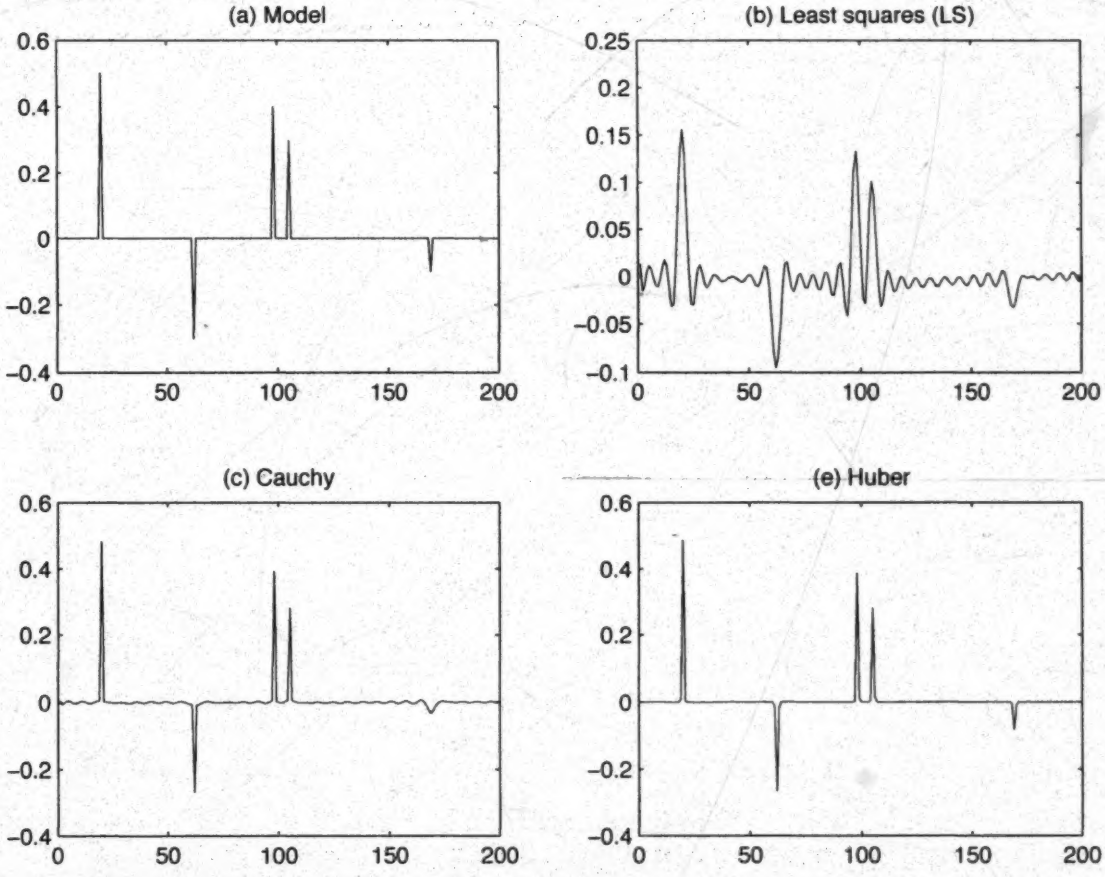

Figure 4.3: Example that compares the least squares solution with the solution obtained via non-quadratic regularization (Huber and Cauchy norms).

For each iteration, the least-squares solution is computed via a conjugate gradient (CG) algorithm explained in the following section. The model weights are updated at every external iteration. Factors that control the convergence of the IRLS solver are the weighting matrix $\mathbf{Q}(\mathbf{m})$, the trade-off parameter μ, and the number of the external and internal iterations (Scales and Gersztenkorn, 1988; Sacchi, 1996; Alliney and Ruzinsky, 1994; Trad, 2001). The weighting function applied in this thesis is estimated by the Cauchy norm. After reducing the cost function (4.5) with Cauchy regularization to the standard form, the hyper-parameter can be set to zero ($\mu = 0$) letting the number of internal iterations play the role of regularizer. The reduction of the cost function to the standard form is described in detail in Section 4.5. and the choice of the hyper-parameter μ is explained in Trad (2001). The external iterations control the sparseness of the solution while trying to fit the data. The convergence is achieved when enough internal and external iterations are used. Recomputing the model weights after more iterations will increase the convergence as opposed to recomputing the weights after every internal iteration. For more information about the convergence of the IRLS algorithm, please refer to Scales and Gersztenkorn (1988), Sacchi (1996), Alliney and Ruzinsky (1994), and Trad (2001).

To summarize, the regularization functions and the solution to the regularized inverse problems along with the desired characteristics of the model are presented in Table 4.1.

4.4 Conjugate gradient algorithm

Conjugate gradient is a semi-iterative method used to solve large systems of equations of type (Hestenes and Steifel, 1952; Scales, 1987; Claerbout, 1992). The system of equations is generally of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4.17}$$

where $\mathbf{A}$ is a known square positive-definite symmetric matrix, $\mathbf{x}$ represents an unknown model vector and $\mathbf{b}$ is a known data vector. The inversion problem presented in the previous sections that we want to solve via CG is of the type

$$\mathbf{L}\mathbf{m} = \mathbf{d} \tag{4.18}$$

Regularization Function $R(\mathbf{m})$	Solution $\mathbf{m}$	Characteristics of the model
$\|\|\mathbf{m}\|\|_2^2$	$(\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}$	Stable/Small
$\|\|\mathbf{D}_1\mathbf{m}\|\|_2^2$	$(\mathbf{L}^T\mathbf{L} + \mu\mathbf{D}_1^T\mathbf{D}_1)^{-1}\mathbf{L}^T\mathbf{d}$	Flat
$\|\|\mathbf{D}_2\mathbf{m}\|\|_2^2$	$(\mathbf{L}^T\mathbf{L} + \mu\mathbf{D}_2^T\mathbf{D}_2)^{-1}\mathbf{L}^T\mathbf{d}$	Smooth
$\sum_{i=1}^{NM} \rho(m_i)$ $\rho(m_i) = \frac{m_i^2}{2}$, if $\|m_i\| \leq a$ $\rho(m_i) = a\|m_i\| - \frac{a^2}{2}$, if $\|m_i\| > a$	$(\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m}))^{-1}\mathbf{L}^T\mathbf{d}$ $Q_{ii} = \|m_i\|$, if $\|m_i\| \leq a$ $Q_{ii} = a[sign(m_i)]$, if $\|m_i\| > a$	Sparse (Huber)
$\sum_{i=1}^{NM} \ln(1 + \frac{m_i^2}{\sigma_m^2})$	$(\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m}))^{-1}\mathbf{L}^T\mathbf{d}$ $Q_{ii} = \frac{1}{1+\frac{m_i^2}{\sigma_m^2}}$	Sparse (Cauchy)

Table 4.1: Regularization functions and their equivalent solutions.

which is a rectangular, under-determined system of equations. This system can be transformed to a different system

$$\mathbf{L}^T\mathbf{L}\mathbf{m} = \mathbf{L}^T\mathbf{d} \tag{4.19}$$

where $\mathbf{L}^T\mathbf{L}$ is a symmetric matrix. If we make the assumption that $\mathbf{L}^T\mathbf{L}$ is also positive-definite then equation (4.18) can be solved using a CG algorithm.The advantage of this algorithm is that the product $\mathbf{L}^T\mathbf{L}$ does not have to be computed. Instead, the operator is solved 'on the fly'. The advantage of the CG over other semi-iterative methods like steepest descent is that it is a much faster technique (Strang, 1986; Claerbout, 1992). There are different implementation of the conjugate gradient method. However, in this thesis I used the conjugate gradient method for least-squares inversion (CGLS) (Scales, 1987). The procedure is as it follows:

Choose an initial solution $\mathbf{m}_0$. A null initial solution ($\mathbf{m}_0 = 0$) is commonly used and gives good results. The data vector $\mathbf{s}$ is then set as $\mathbf{s}_0 = \mathbf{d} - \mathbf{L}\mathbf{m}_0$. Two complementary model vectors $\mathbf{r}$ and $\mathbf{p}$ are initialized: $\mathbf{r}_0 = \mathbf{p}_0 = \mathbf{L}^T(\mathbf{d} - \mathbf{L}\mathbf{m}_0)$ and the second data vector $\mathbf{q}$ is computed $\mathbf{q}_0 = \mathbf{L}\mathbf{p}_0$.

Then for $k = 0, 1, ...$:

$\alpha_{k+1} = \frac{(\mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{q}_k, \mathbf{q}_k)}$

$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_{k+1}\mathbf{p}_k$

$\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_{k+1}\mathbf{q}_k$

$\mathbf{r}_{k+1} = \mathbf{L}^T\mathbf{s}_{k+1}$

$\beta_{k+1} = \frac{(\mathbf{r}_{k+1}, \mathbf{r}_{k+1})}{(\mathbf{r}_k, \mathbf{r}_k)}$

$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_{k+1}\mathbf{p}_k$

and

$\mathbf{q}_{k+1} = \mathbf{L}\mathbf{p}_{k+1}$,

where $(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T\mathbf{x} = \mathbf{x} \cdot \mathbf{x}$ and $\mathbf{L}$ and $\mathbf{L}^T$ are the forward and adjoint Radon operator, respectively.

If we want to include additional operators such as regularization operators the system of equations (4.17) can be modified and written as an augmented system

$$\begin{pmatrix} \mathbf{L} \\ \mu \mathbf{W}_m \end{pmatrix} \mathbf{m} = \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix} \tag{4.20}$$

The above system will be solved minimizing the cost function

$$J = ||\mathbf{L}\mathbf{m} - \mathbf{d}||_2^2 + \mu ||\mathbf{W}_m \mathbf{m}||_2^2 \tag{4.21}$$

which can be reduced to the standard form (Hanke and Hansen, 1993) making it easier to incorporate in the conjugate gradient (CG) algorithm.

4.5 Reduction to standard form and Preconditioning

In order to improve the computational efficiency of regularized inverse problem the system of normal equations (4.19) can be reduced to the standard form (4.1) (Hanke and Hansen, 1993). The standard form of a cost function is considered to be of the type

$$J = ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + \mu ||\mathbf{x}||_2^2. \tag{4.22}$$

If we consider $\mathbf{W}_m \mathbf{m} = \tilde{\mathbf{m}}$ in equation (4.20) then $\mathbf{m} = \mathbf{W}_m^{-1} \tilde{\mathbf{m}}$ and the cost function becomes

$$J = ||\mathbf{L}\mathbf{W}_m^{-1} \tilde{\mathbf{m}} - \mathbf{d}||_2^2 + \mu ||\tilde{\mathbf{m}}||_2^2. \tag{4.23}$$

If $\mathbf{L}\mathbf{W}_\mathbf{m}^{-1} = \tilde{\mathbf{L}}$ and $\mathbf{d} = \tilde{\mathbf{d}}$ then the above objective function can be written as

$$J = ||\tilde{\mathbf{L}}\tilde{\mathbf{m}} - \tilde{\mathbf{d}}||_2^2 + \mu ||\tilde{\mathbf{m}}||_2^2 \tag{4.24}$$

which is similar to the standard form defined in (4.21) (Hanke and Hansen, 1993). A CGLS algorithm can be applied to compute the model $\tilde{\mathbf{m}}$

$$\tilde{\mathbf{m}} = (\tilde{\mathbf{L}}^T \tilde{\mathbf{L}} + \mu \mathbf{I})^{-1} \tilde{\mathbf{L}}^T \tilde{\mathbf{d}}. \tag{4.25}$$

To recover the solution of interest $\mathbf{m}$ we can now just multiply equation (4.24) by $\mathbf{W}_m^{-1}$

$$\mathbf{m} = \mathbf{W}_m^{-1} \tilde{\mathbf{m}} = \mathbf{W}_m^{-1} (\tilde{\mathbf{L}}^T \tilde{\mathbf{L}} + \mu \mathbf{I})^{-1} \tilde{\mathbf{L}}^T \tilde{\mathbf{d}}. \tag{4.26}$$

This is equivalent to

$$\mathbf{m} = \mathbf{W}_m^{-1}((\mathbf{L}\mathbf{W}_m^{-1})^T(\mathbf{L}\mathbf{W}_m^{-1}) + \mu\mathbf{I})^{-1}(\mathbf{L}\mathbf{W}_m^{-1})^T\mathbf{d}, \tag{4.27}$$

which can be written as

$$\mathbf{m} = \mathbf{W}_m^{-1}(\mathbf{W}_m^{-T}\mathbf{L}^T\mathbf{L}\mathbf{W}_m^{-1} + \mu\mathbf{I})^{-1}\mathbf{W}_m^{-T}\mathbf{L}^T\mathbf{d}, \tag{4.28}$$

and

$$\mathbf{m} = (\mathbf{W}_m\mathbf{W}_m^{-T}\mathbf{L}^T\mathbf{L}\mathbf{W}_m^{-1}\mathbf{W}_m^T + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}. \tag{4.29}$$

If the weighting matrix $\mathbf{W}_m$ and its corresponding inverse $\mathbf{W}_m^{-1}$ are diagonal matrices, like in the case of sparseness constraints, then $\mathbf{W}_m^T = \mathbf{W}_m$ and $\mathbf{W}_m^{-T} = \mathbf{W}_m^{-1}$, therefore equation (4.28) reduces to

$$\mathbf{m} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{W}_m^T\mathbf{W}_m)^{-1}\mathbf{L}^T\mathbf{d}, \tag{4.30}$$

exactly the solution presented in (4.7). If the weighting matrix is not diagonal, like in the case of the first and second derivative matrices used as smoothness constraints, the reduction to the standard form becomes more complicated and $\mathbf{W}_m^{-1}$ is replaced by the pseudo-inverse $\mathbf{W}_m^{\dagger}$ (Hanke and Hansen, 1993; Trad, 2001). In this thesis I use Cauchy regularization norms which give rise to diagonal weighting matrices thus making the transformation to the standard form rather trivial.

By reducing our cost function to the standard form we now try to solve the following inverse problem

$$\tilde{\mathbf{L}}\tilde{\mathbf{m}} = \tilde{\mathbf{d}} \tag{4.31}$$

which is equivalent to

$$\mathbf{L}\mathbf{W}_m^{-1}\tilde{\mathbf{m}} = \tilde{\mathbf{d}}. \tag{4.32}$$

This is similar to right preconditioning the system of equations (4.17) with the preconditioner $\mathbf{W}_m^{-1}$. Preconditioning techniques are generally used in iterative methods like

CG to modify the structure of a matrix and accelerate the convergence rate of the method which depends on the spectral characteristics of the operator (Saad, 1996; Alleon et al., 1997; Benzi, 2002). In the case of a discrete ill-posed inverse problem (i.e. geophysical inverse problems) the preconditioner is rather used to improve the quality of the solution than speed up the convergence rate (Calvetti, 2006). The right side preconditioning is generally linked to the *a priori* information about the solution (i.e. sparseness in this case) while the left side preconditioning is connected with the noise in the data (Calvetti, 2006). The general definition of right-preconditioning states that it transforms a linear system of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4.33}$$

to a linear system of the form

$$\mathbf{A}\mathbf{M}^{-1}\mathbf{u} = \mathbf{b}, \tag{4.34}$$

where $\mathbf{x} = \mathbf{M}^{-1}\mathbf{u}$. It can be noticed that systems (4.31) and (4.33) are the same, with $\mathbf{A} = \mathbf{L}$, $\mathbf{M}^{-1} = \mathbf{W}_m^{-1}$, $\mathbf{x} = \mathbf{m}$, $\mathbf{u} = \tilde{\mathbf{m}}$, and $\mathbf{b} = \tilde{\mathbf{d}} = \mathbf{d}$.

The implementation of the right preconditioned CGLS algorithm (Hanke, 1995; Hansen, 1998) is at it follows:

Choose an initial solution $\mathbf{m}_0$ ($\mathbf{m}_0 = 0$ is generally used). The data vector $\mathbf{s}$ is computed as $\mathbf{s}_0 = \mathbf{d} - \mathbf{L}\mathbf{m}_0 = \mathbf{d} - \mathbf{L}\mathbf{W}_m^{-1}\tilde{\mathbf{m}}_0$. The new variable $\tilde{\mathbf{m}}$ does not need to be initially set. The algorithm needs $\mathbf{m}_0$ as an input, although the final solution is given in terms of $\tilde{\mathbf{m}}$. The auxiliary model vectors $\mathbf{r}$ and $\mathbf{p}$ are set as $\mathbf{r}_0 = \mathbf{p}_0 = \mathbf{W}_m^{-1}\mathbf{L}^T\mathbf{s}_0$ and the second data vector $\mathbf{q}$ is computed $\mathbf{q}_0 = \mathbf{L}\mathbf{p}_0$.

Then for $k = 0, 1, \ldots$:

$\alpha_{k+1} = \frac{(\mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{q}_k, \mathbf{q}_k)}$

$\tilde{\mathbf{m}}_{k+1} = \mathbf{m}_k + \alpha_{k+1}\mathbf{p}_k$

$\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_{k+1}\mathbf{q}_k$

$\mathbf{r}_{k+1} = \mathbf{W}_m^{-1}\mathbf{L}^T\mathbf{s}_{k+1}$

$\beta_{k+1} = \frac{(\mathbf{r}_{k+1}, \mathbf{r}_{k+1})}{(\mathbf{r}_k, \mathbf{r}_k)}$

$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_{k+1}\mathbf{p}_k$

and

$\mathbf{q}_{k+1} = \mathbf{L}\mathbf{W}_m^{-1}\mathbf{p}_{k+1}$,

where $(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T\mathbf{x} = \mathbf{x} \cdot \mathbf{x}$ and $\mathbf{L}\mathbf{W}_m^{-1}$ and $\mathbf{W}_m^{-1}\mathbf{L}^T$ are the forward and adjoint Radon operator, respectively. After the CGLS algorithm is stopped and the final preconditioned solution $\tilde{\mathbf{m}}$ is computed, by multiplying it by $\mathbf{W}_m^{-1}$ we will obtain the desired solution $\mathbf{m}$.

The algorithm can be stopped at any step and the intermediate solution analyzed. The stopping criterion used in this thesis is a variant of the Generalized Cross Validation Criterion (GCV) which depends on the norm of the residual $||\mathbf{r}_{k+1}||_2^2$ (Haber, 1997)

$$GCV(iter) = \frac{\sum_{k=1}^{N} ||\mathbf{r}_{k+1}||_2^2}{N - iter^2} \tag{4.35}$$

where N is the number of iterations previously set. The algorithm is stopped either when the number of iterations *iter* reaches the maximum number of iterations previously set ($iter = N$) or when the data residual is small enough. The threshold for the data residual is considered to be the standard deviation of the random noise. This value is generally not known in seismic data, therefore a threshold value is generally considered.

There are other preconditioning methods such as left preconditioning, and split preconditioning, however, they are not the subject of this thesis, therefore they will not be discussed. For more information about them, please refer to Saad (1996) and/or Benzi (2002).

4.6 Summary

In this chapter I introduced the linear inversion method as a mean to solve a system of normal equations. A classification of the inverse problems in terms of the number of knowns and unknowns has been presented. Most of the inverse problems encountered in geophysics are ill-posed, therefore it is needed to introduce regularization functions in the computation of the solution. Regularization functions are often presented as

weighting functions that contain *a priori* information about the solution. Emphasis has been placed on sparseness constraints and the corresponding regularization norms. For large and sparse inverse problems iterative methods like conjugate gradients (CG) are preferred. The method used in this thesis is conjugate gradient least squares (CGLS) introduced by Scales (1987). To incorporate regularization terms in the CGLS algorithm the system (4.19) is reduced to the standard form which is equivalent to applying right preconditioning to the initial system (4.1). The right preconditioned CGLS method is also described.

Chapter 5

Hybrid Radon Transform

5.1 Introduction

Surface waves, known as ground roll have low-velocity, low-frequency, and often exhibit high energy that can obscure signal and deteriorate the quality of the data. In seismic processing we often look to transform the data to a new domain where signal and noise are readily separable therefore filtering methods can be applied. For example, when a 2-D Fourier transform is applied to a shot gather an $f-k$ spectrum is obtained. In the $f-k$ domain all the events are transformed into linear events. Every linear event in the frequency-wavenumber domain is equivalent to events with the same dip in the time-offset domain which means that events with different dips that intersect in the $t-h$ domain can be separated in the $f-k$ domain. The main problem in $f-k$ filtering is spatial aliasing which causes events with steep dips to map to higher frequencies than they would normally map. In the case of ground roll suppression, spectral components of the noise will overlap spectral components of the signal and either the noise will not be completely eliminated or the signal will be distorted. Radon transform (RT) has properties that make it effective for coherent noise attenuation. As previously shown, linear Radon transform (LRT) would focus linear events and discriminate them based on the ray parameter, while the hyperbolic events would be focused by the hyperbolic Radon transform and discriminated based on velocity. A linear RT applied to hyperbolas will transform them into ellipses in Radon domain (Fig. 5.1(b)), and a hyperbolic RT applied to lines will transform them into higher order degree lines (Fig. 5.1(c)). In the case of seismic data that contains primary reflections and ground roll the use of only a linear Radon

transform (LRT) or a hyperbolic Radon transform (HRT) will not focus both events at the same time. One solution is to adopt a combined operator that contains both hyperbolic and linear integration paths and simultaneously model the signal and noise. This Radon transform was first introduced in Trad et. al (2001) and was denominated the *Hybrid Radon Transform*. A similar approach was introduced by Chen et al. (1998) to decompose a signal into an optimal superposition of dictionary elements. The procedure is called *basis pursuit* and the optimization is done by imposing sparseness constraints given by an l_1 norm for the model.

5.2 Theory

We can express the seismic data $\mathbf{d}$ as a sum of signal $\mathbf{d}_h$, coherent noise (linear events) $\mathbf{d}_l$ and additive noise $\mathbf{n}$

$$\mathbf{d} = \mathbf{d}_h + \mathbf{d}_l + \mathbf{n} \tag{5.1}$$

We will model the signal using the hyperbolic Radon transform ($\mathbf{L}_h$) and the ground roll using the linear Radon transform ($\mathbf{L}_l$), in which case they can be described as

$$\mathbf{d}_h = \mathbf{L}_h \mathbf{m}_h \tag{5.2}$$

and

$$\mathbf{d}_l = \mathbf{L}_l \mathbf{m}_l \tag{5.3}$$

respectively, where $\mathbf{m}_h$ denotes the hyperbolic Radon panel and $\mathbf{m}_l$ represents the linear Radon panel.

The linear and hyperbolic Radon transforms were defined in Chapter 3. The two modeling operators $\mathbf{L}_h$ and $\mathbf{L}_l$ form a combined operator $\mathbf{L}$ that acts on the two models as

$$\mathbf{d} = \mathbf{L}\mathbf{m} = (\mathbf{L}_h \; \mathbf{L}_l)\begin{pmatrix}\mathbf{m}_h \\ \mathbf{m}_l\end{pmatrix} = \mathbf{L}_h\mathbf{m}_h + \mathbf{L}_l\mathbf{m}_l \tag{5.4}$$

where $\mathbf{m}$ is a combination of the two models $\mathbf{m}_h$ and $\mathbf{m}_l$.

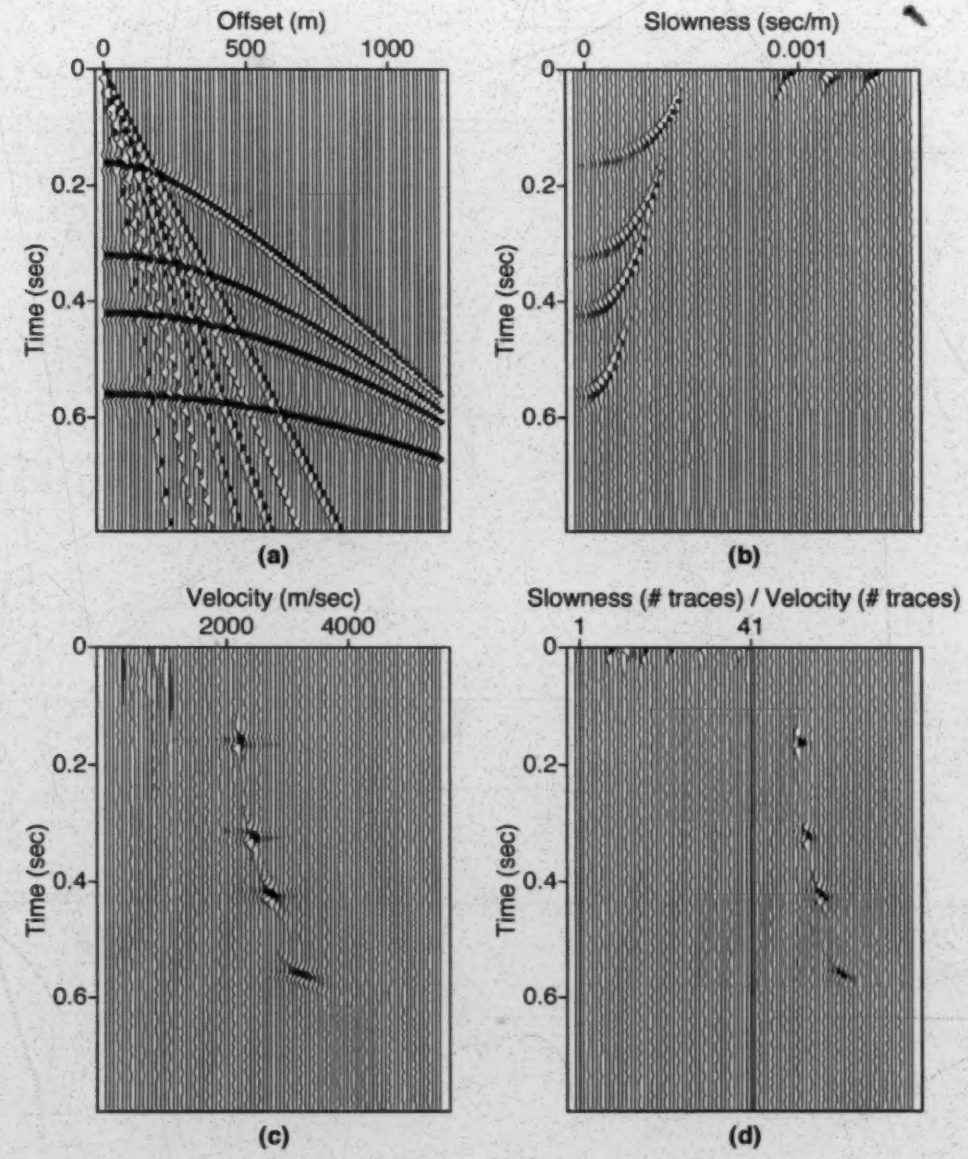

Figure 5.1: Example illustrating the result of applying different Radon transforms to synthetic data containing hyperbolic and linear events. (a) Synthetic data containing four hyperbolic events and seven linear events. (b) Previous data in the linear Radon domain. Lines are focused while the hyperbolas are transformed into ellipses. (c) Same data in the hyperbolic Radon domain. Hyperbolas are focused and lines are not. (d) Synthetic data in the hybrid Radon domain. Both kinds of events are now focused.

The standard approach to ground roll elimination is to estimate the noise component and subtract it from the original data, thus obtaining the primary reflections. By using the hybrid Radon transform we simultaneously invert for the Radon panels $\mathbf{m}_h$ and $\mathbf{m}_l$.

As shown before, this problem can be solved by means of inversion. Giving the following objective function

$$J = \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \mu R(\mathbf{m}) \tag{5.5}$$

where μ is the trade-off parameter and

$$R(\mathbf{m}) = \sum_{i=1}^{NM} \ln\left(1 + \frac{m_i^2}{\sigma_m^2}\right) \tag{5.6}$$

is the Cauchy regularization norm, and minimizing it with respect to $\mathbf{m}$ we obtain the least squares (LS) solution

$$\hat{\mathbf{m}} = \left(\mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m})\right)^{-1}\mathbf{L}^T\mathbf{d} \tag{5.7}$$

which is equivalent to

$$\begin{pmatrix}\hat{\mathbf{m}}_h \\ \hat{\mathbf{m}}_l\end{pmatrix} = \left(\mathbf{L}_h^T\mathbf{L}_h + \mathbf{L}_l^T\mathbf{L}_l + \mu\mathbf{Q}(\mathbf{m})\right)^{-1}\begin{pmatrix}\mathbf{L}_h^T \\ \mathbf{L}_l^T\end{pmatrix}\mathbf{d} \tag{5.8}$$

where $\mathbf{Q}(\mathbf{m})$ is a diagonal matrix whose elements are equal to

$$Q_{ii} = \frac{1}{1 + \frac{m_i^2}{\sigma_m^2}} = \frac{\sigma_m^2}{\sigma_m^2 + m_i^2} \tag{5.9}$$

If we consider $\lambda = \mu\sigma_m^2$, then the solution becomes

$$\begin{pmatrix}\hat{\mathbf{m}}_h \\ \hat{\mathbf{m}}_l\end{pmatrix} = \left(\mathbf{L}_h^T\mathbf{L}_h + \mathbf{L}_l{}^T\mathbf{L}_l + \lambda\mathbf{Q}'(\mathbf{m})\right)^{-1}\begin{pmatrix}\mathbf{L}_h^T \\ \mathbf{L}_l^T\end{pmatrix}\mathbf{d} \tag{5.10}$$

where $\mathbf{Q}'(\mathbf{m})$ is a diagonal matrix whose elements are equal to

$$Q'_{ii} = \frac{1}{\sigma_m^2 + m_i^2} \tag{5.11}$$

The parameter σ_m acts as a hyper-parameter that allows a trade-off between the degree of data fit and sparseness of the solution (Thorson and Claerbout, 1985; Sacchi and Ulrych,

1995). Choosing an appropriate parameter σ_m can be difficult but in time domain the solution is not very sensitive to σ_m and there is need just for one value. This value is empirically set to some percentage of the maximum amplitude of the previous solution. The final solution can be found using a right preconditioned iterative reweighted least squares conjugate gradient (IRLS) method described in the previous chapter.

The estimated signal or noise can now be calculated using only one of the operators:

$$\hat{\mathbf{d}}_h = \mathbf{L}_h \hat{\mathbf{m}}_h, \tag{5.12}$$

and

$$\hat{\mathbf{d}}_l = \mathbf{L}_l \hat{\mathbf{m}}_l. \tag{5.13}$$

We obtain a good result as long as the basis functions of constitutive Radon transforms are different from each other. If they resemble in any way, the same event can be mapped in both spaces making the separation of events more difficult. For this case, at long offset the hyperbolas asymptotically resemble lines, therefore, energy belonging to the reflections would be mapped in the linear Radon panel and interpreted as ground roll energy. This phenomenon of contaminating signals by other signals is commonly known as crosstalk (Claerbout, 1992).

We could overcome the crosstalk problem by including extra terms in the regularization function. If we introduce the regularization terms as *a priori* probability density functions and explain the inversion in Bayesian terms (Menke, 1984; Ulrych et al., 2001; Trad, 2001; Youzwishen, 2001;) we can link the model weighting functions $\mathbf{Q}(\mathbf{m})$ to the model covariance matrix $\mathbf{C}_m$ that characterizes the basic shape of joint distribution ($\mathbf{Q}(\mathbf{m}) = \mathbf{C}_m^{-1}$). In the hybrid Radon case, we can write the model covariance matrix as

$$\mathbf{C}_m = \begin{pmatrix} \mathbf{C}_{ll} & \mathbf{C}_{lh} \\ \mathbf{C}_{hl} & \mathbf{C}_{hh} \end{pmatrix}$$

where the diagonal elements are a measure of the width of the distribution of the model, and the off-diagonal elements indicate the degree to which the pairs of model are correlated. To translate this to the hybrid Radon transform, when the basis functions that describe the signal (reflections) and noise (ground roll) are different, we consider the

models to be uncorrelated $\mathbf{C}_{hl} = \mathbf{C}_{lh} = 0$. This is equivalent to the case presented at the beginning of this chapter. If there is crosstalk then the models $\mathbf{m}_h$ and $\mathbf{m}_l$ are correlated and $\mathbf{C}_{hl} \neq \mathbf{C}_{lh} \neq 0$. The degree of crosstalk should determine how much weight is accorded to which off-diagonal term. However, this problem is more difficult to implement as a preconditioned iterative reweighted least squares (IRLS) algorithm and is not the subject of this thesis.

5.3 Synthetic data examples

The first synthetic data example illustrates the advantages of a hybrid Radon transform. Four hyperbolic reflections with velocities between 2250 m/sec and 3250 m/sec intersect six linear events with velocities ranging from 320 m/sec to 600 m/sec (Fig. 5.1(a)). Figure 5.1(c) shows the Radon panel when only the hyperbolic Radon transform is applied to the data. When the hybrid Radon transform is applied (Fig. 5.1(d)) the hyperbolic Radon operator maps the hyperbolic events to the second area of the model (traces 41 and up), while the linear Radon operator maps the linear events to the first area of the model (traces 1 to 40). Using a conjugate gradient method we simultaneously invert for the two models $\mathbf{m_h}$ and $\mathbf{m_l}$. Figure 5.2(a) shows the result after five iterations with quadratic regularization (damped least squares solution), and Figure 5.2(b) shows the model after two external iterations and five internal iterations each time, when the Cauchy regularization described in equation (5.6) is applied (high resolution solution). It can be observed that by including a sparseness constraint in the inversion algorithm a more focused solution is obtained.

One can now apply the forward hyperbolic Radon operator to the hyperbolic models in Figure 5.2 (equation (5.9)) and obtain only the clean reflection data (Fig. 5.3(a) and Fig. 5.3(b)) or apply the forward linear Radon operator to the linear models in the same figure (equation 5.10) and obtain the coherent noise, which can be subtracted from the original data to obtain the noise free data. Figures 5.3(c) and 5.3(d) show the difference between the original data (Fig. 5.1(a))and the recovered hyperbolic reflections (Fig. 5.3(a) and 5.3(b)). It can be noticed that the hyperbolic data is better recovered from the high resolution hybrid Radon panel than from the damped least squares hybrid Radon panel.

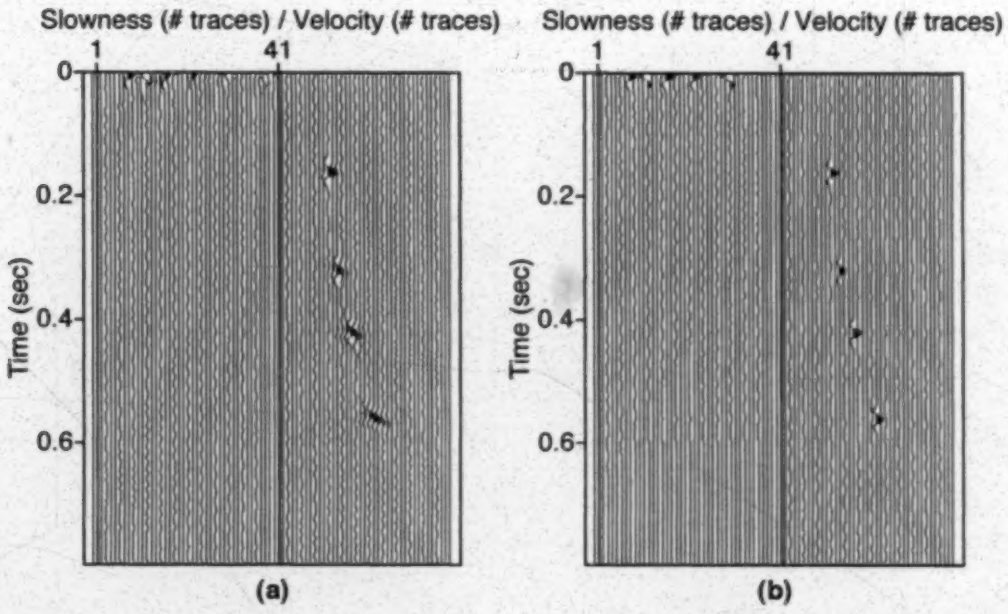

Figure 5.2: Synthetic data example comparing the damped least squares (DLS) and the high resolution solutions for hybrid Radon transform. (a) Damped least squares solution obtained after 5 iterations of a conjugate gradient least squares (CGLS) algorithm. (b) High resolution solution obtained after 2 external iterations and 5 internal iterations each time when an iterative reweighted least squares (IRLS) conjugate gradient algorithm was applied. A sparseness constraint of the model is introduced via a Cauchy regularization norm. It can be observed that the high resolution solution is more focused than the damped least squares solution.

Figure 5.4 strengthens this conclusion by comparing the predicted data from the two models (Fig. 5.4(a) and 5.4(b)) and the difference between the original and the predicted data (Fig.5.4(c) and 5.4(d)). As it can be seen in Figures 5.4(b) and (d) the data are almost entirely reconstructed except the two linear events with the larger slope. The reason for not recovering these two events is aliasing in the $\tau - p$ transformation (Maroof, 1984).

The second synthetic data example illustrates the performance of hybrid Radon transform on noisy data. I use the same synthetic data shown in Figure 5.1(a) but in this case I added Gaussian random noise (Figure 5.5(a)). The signal to noise ratio I used is 6 (SNR=6). In Figure 5.5 I present the damped least squares solution after five iterations

(Fig. 5.5(b)) and the recovered data (Fig. 5.5(c)) along with the difference between the original and the recovered data (Fig. 5.5(d)). As it can be observed in Figure 5.5, the DLS recovers not only the signal and the coherent noise, but also part of the additive noise.

Figure 5.6 shows the IRLS solution after two external iterations and five internal iterations. The random noise is suppressed but the linear and hyperbolic events are not entirely recovered either (Fig. 5.6(c)).

Figure 5.7 shows the IRLS solution after two external iterations and twenty internal iterations. The predicted data show that the random noise and artifacts were eliminated due to the sparseness in Radon domain but at the same time, by increasing the number of internal iterations a larger number of the hyperbolic and linear events were recovered (Fig. 5.7(c)). A conclusion to be drawn from Figures 5.5, 5.6 and 5.7 is that the external iterations control the sparseness in the Radon domain, while the internal iterations are responsible for the prediction of the data.

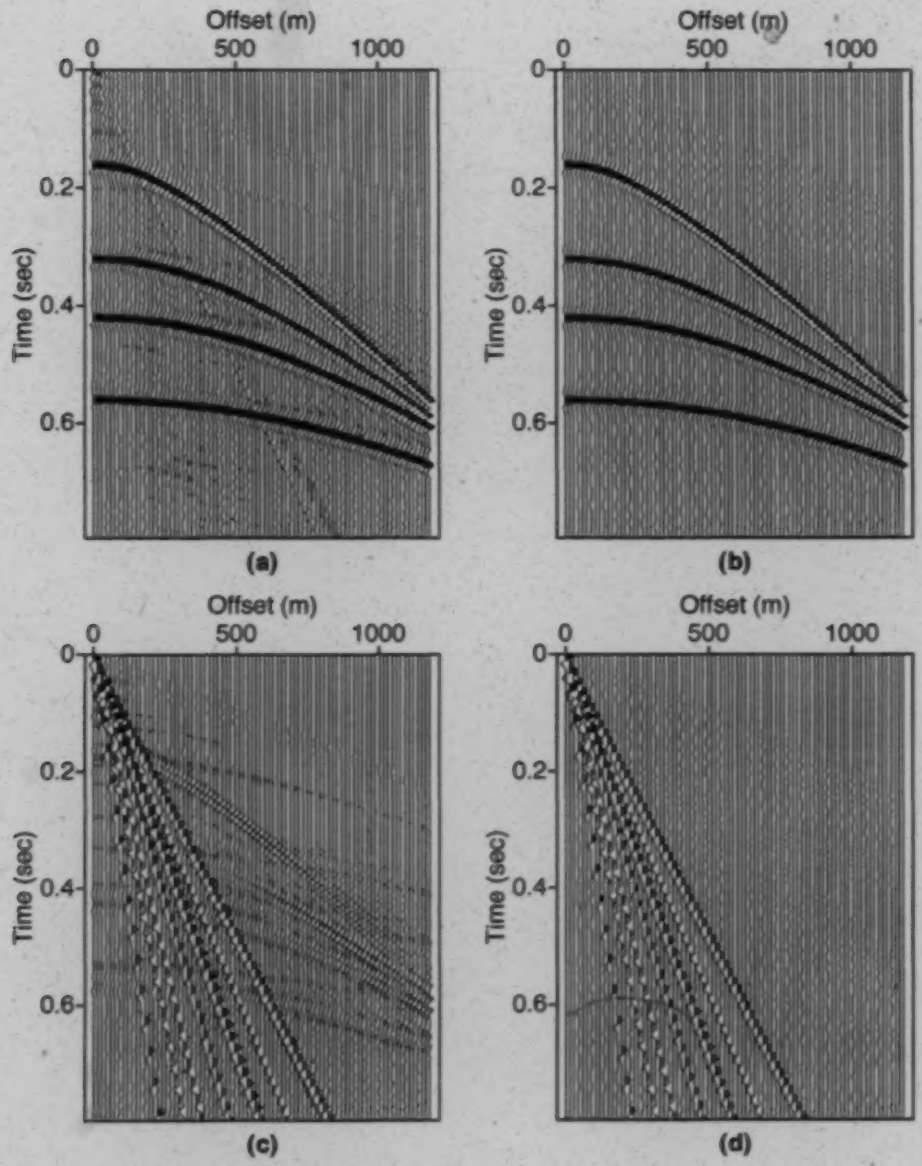

Figure 5.3: Synthetic data illustrating the difference between the damped least squares solution (DLS) and the iterative reweighted least squares (IRLS) solution. (a) Hyperbolic data recovered from the DLS panel. (b) Hyperbolic data recovered from the IRLS panel. (c) The difference between the original data (Fig. 5.1(a)) and the predicted hyperbolas from (a). (d) The difference between the original data (Fig. 5.1(a)) and the predicted hyperbolas from (b).

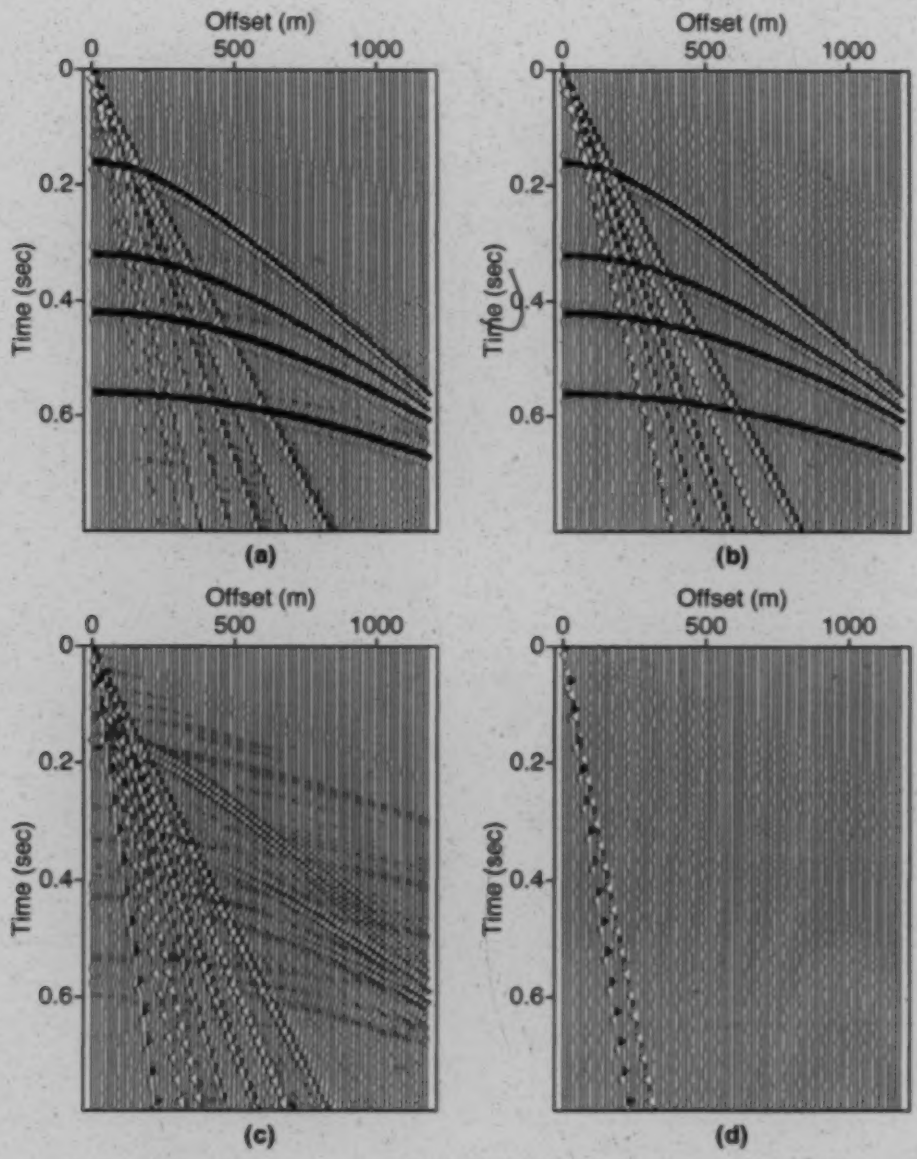

Figure 5.4: Synthetic data comparing the data prediction from the damped least squares solution (DLS) and the iterative reweighted least squares (IRLS) solution. (a) Data predicted from the DLS panel. (b) Data predicted from the IRLS panel. (c) The difference between the original data (Fig. 5.1(a)) and the predicted data from (a). (d) The difference between the original data (Fig. 5.1(a)) and the predicted data from (b).

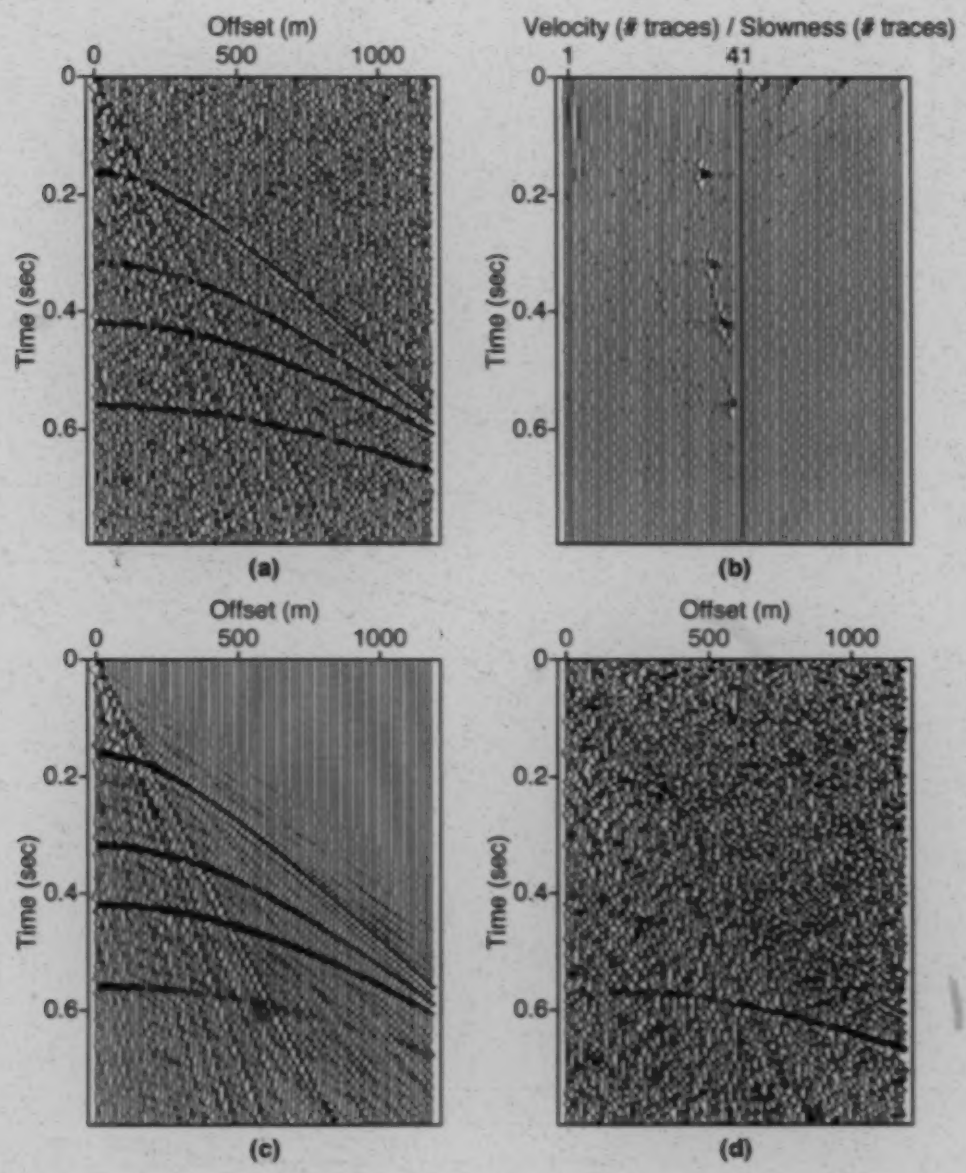

Figure 5.5: Synthetic data example illustrating the performance of the hybrid Radon transform on noisy data. (a) Noisy synthetic data. Signal to noise ratio is 6 (SNR=6). (b) DLS hybrid Radon panel after 5 iterations. (c) The recovered data from the DLS hybrid Radon panel. (d) The difference between the original data (a) and the predicted data (c).

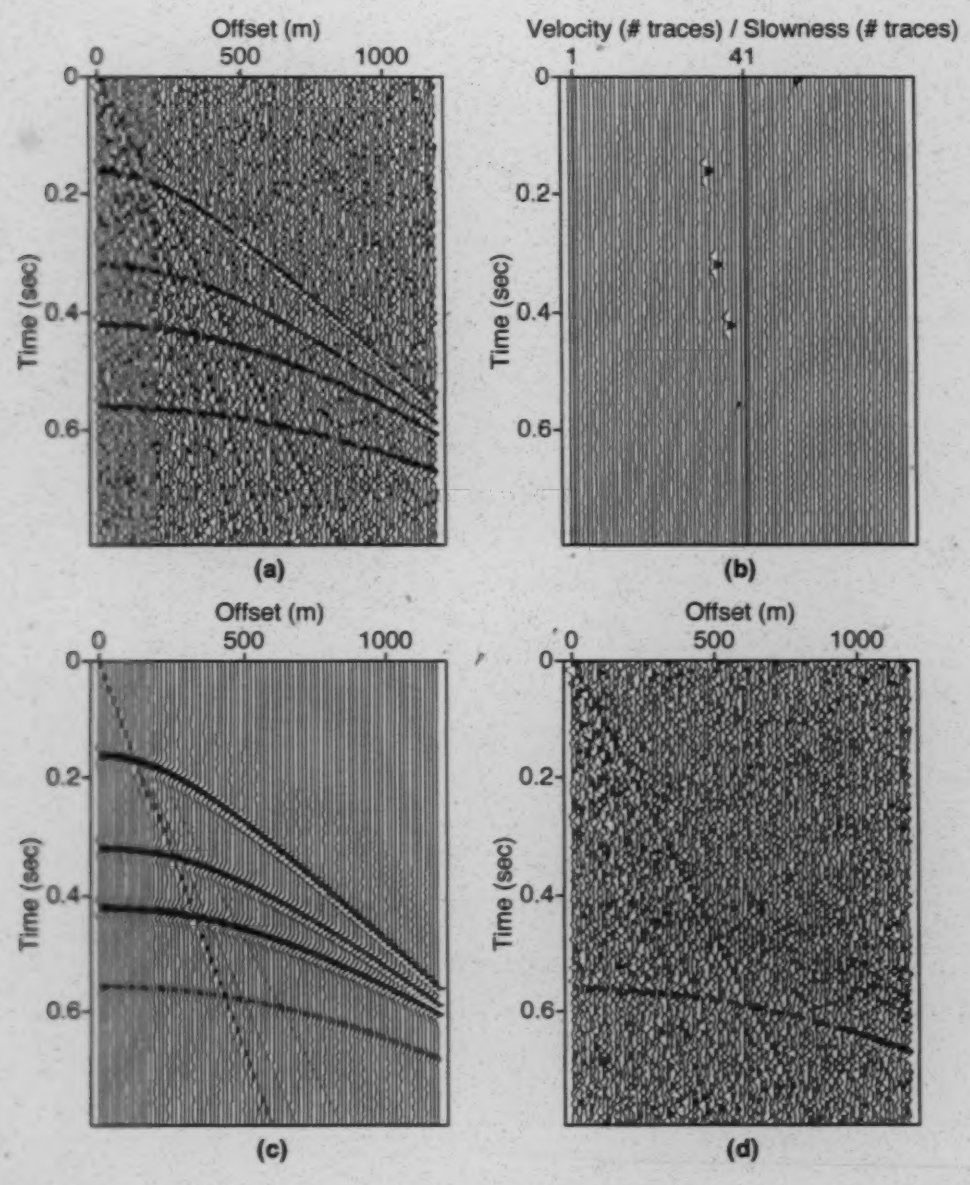

Figure 5.6: Synthetic data example illustrating the performance of the hybrid Radon transform on noisy data. (a) Noisy synthetic data. Signal to noise ratio is 6 (SNR=6). (b) IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations (c) The recovered data from the DLS hybrid Radon panel. (d) The difference between the original data (a) and the predicted data from (c).

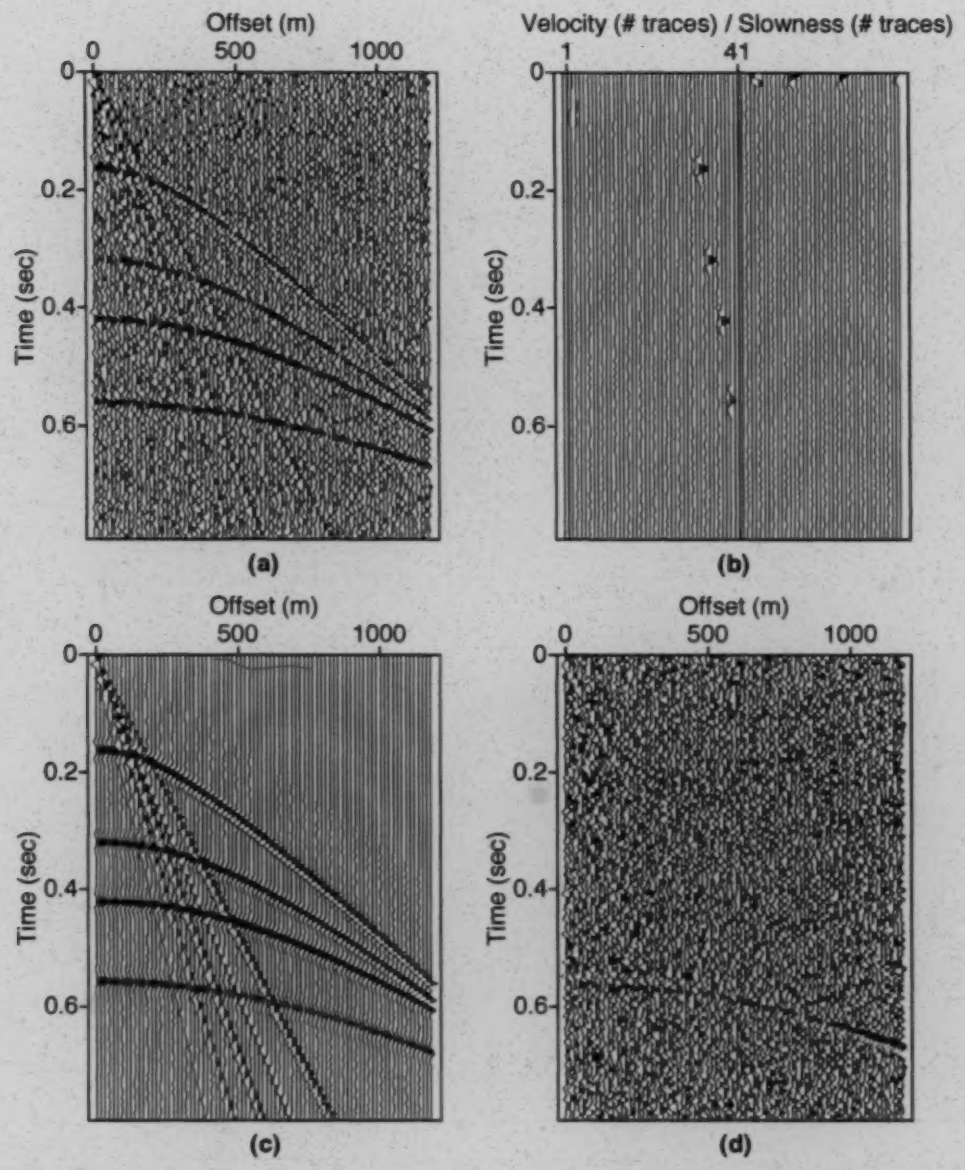

Figure 5.7: Synthetic data example illustrating the performance of the hybrid Radon transform on noisy data. (a) Noisy synthetic data. Signal to noise ratio is 6 (SNR=6). (b) IRLS hybrid Radon panel after 2 external iterations and 20 internal iterations (c) The recovered data from the DLS hybrid Radon panel. (d) The difference between the original data (a) and the predicted data from (c).

5.4 Real data examples

The first real data example illustrates the difference between the damped least squares solution (DLS) and the high resolution iterative reweighted least squares solution (IRLS), as well as the influence of the hyper-parameter σ_m^2 on the results. Figure 5.8(a) shows a field data record that contains reflections, ground roll, direct arrival and random noise. Figure 5.8(b) represents the DLS hybrid Radon panel after five iterations and Figure 5.8(c) presents the recovered data from the DLS solution. Figure 5.9 shows the original data (Fig. 5.9(a)), the recovered reflections (Fig. 5.9(b)), the recovered ground roll (Fig. 5.9(c)) and the ground roll free data (Fig. 5.9(d)) which in this case represents the predicted ground roll subtracted from the data. This is preferable to keeping the recovered hyperbolic data, because part of the signal may be lost when the forward Radon transform is applied.

Figure 5.10 compares the DLS solution after five (Fig. 5.10(a)) and thirty (Fig. 5.10(b)) iterations with the IRLS solution after two external iterations and five (Fig.5.10(c)) and thirty (Fig. 5.10(d)) internal iterations, respectively. The parameter σ_m^2 used for the IRLS solution is set equal to ten percent of the DLS solution with maximum amplitude ($\sigma_m^2 = 0.1max|m_{DLS}|$).

Figure 5.11 compares the DLS solution after five and thirty iterations with the IRLS solution after two external iterations and five (Fig.5.11(c)) and thirty (Fig. 5.11(d)) internal iterations, respectively, only this time the hyper-parameter σ_m^2 is set as fifty percent of the DLS solution with maximum amplitude ($\sigma_m^2 = 0.5max|m_{DLS}|$).

To describe the role of the external iterations, lets take a look at Figures 5.10(a) and (c). Figure 5.10(a) represents the DLS solution after five iterations which is equivalent to applying the IRLS algorithm using one external iteration and five internal iterations. Figure 5.10 (c), as mentioned above, represents the IRLS solution after two external iterations and five internal iterations. As it can be observed, the Radon panel in Figure 5.10 (c) is more sparse than the Radon panel in Figure 5.10(a) helping us to draw the conclusion that the number of external iterations controls the sparseness of the solution. The same can be observed from comparing Figure 5.10(b) with (d), Figure 5.11(a) with (c), and Figure 5.11(b) with (d).

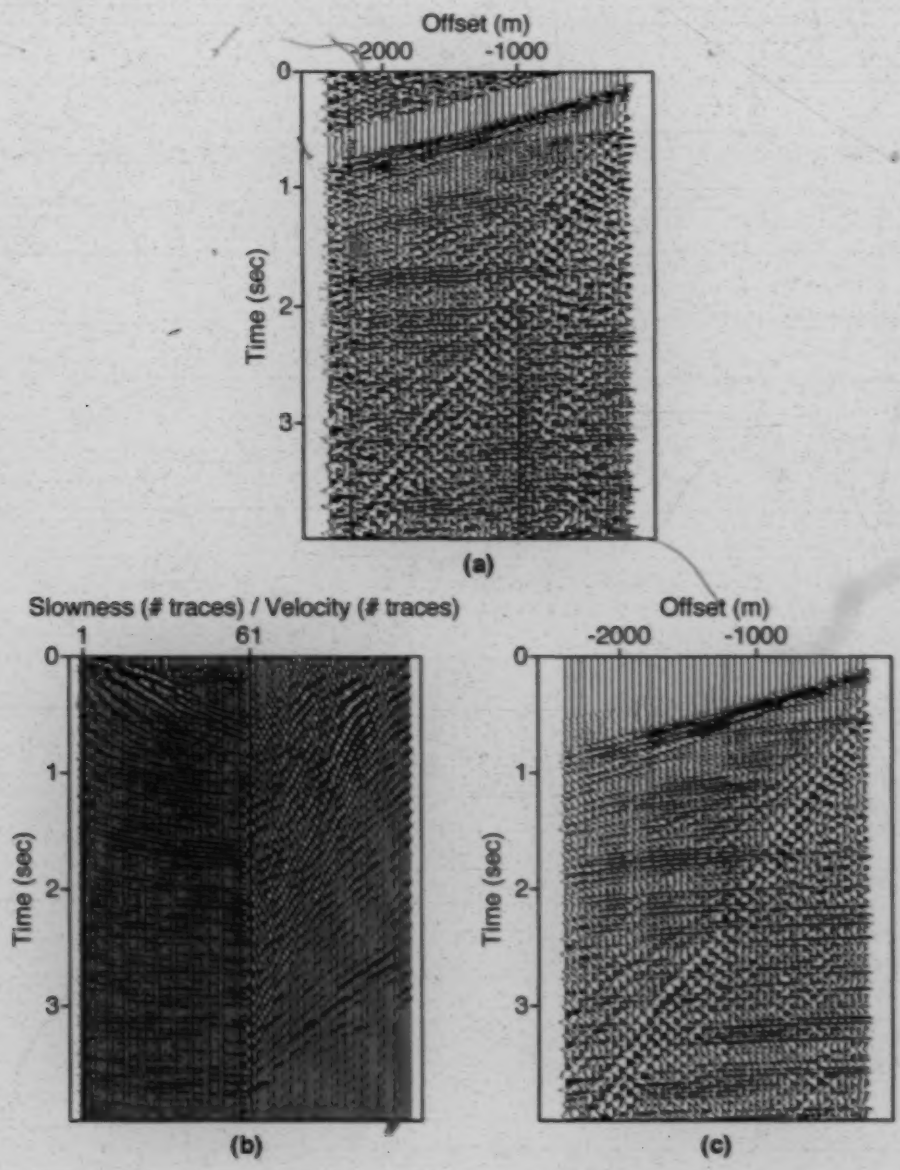

Figure 5.8: (a) Field data record containing reflections, ground roll, direct arrival, and random noise. (b) DLS hybrid Radon panel after 5 iterations. (c) Data recovered from the DLS panel.

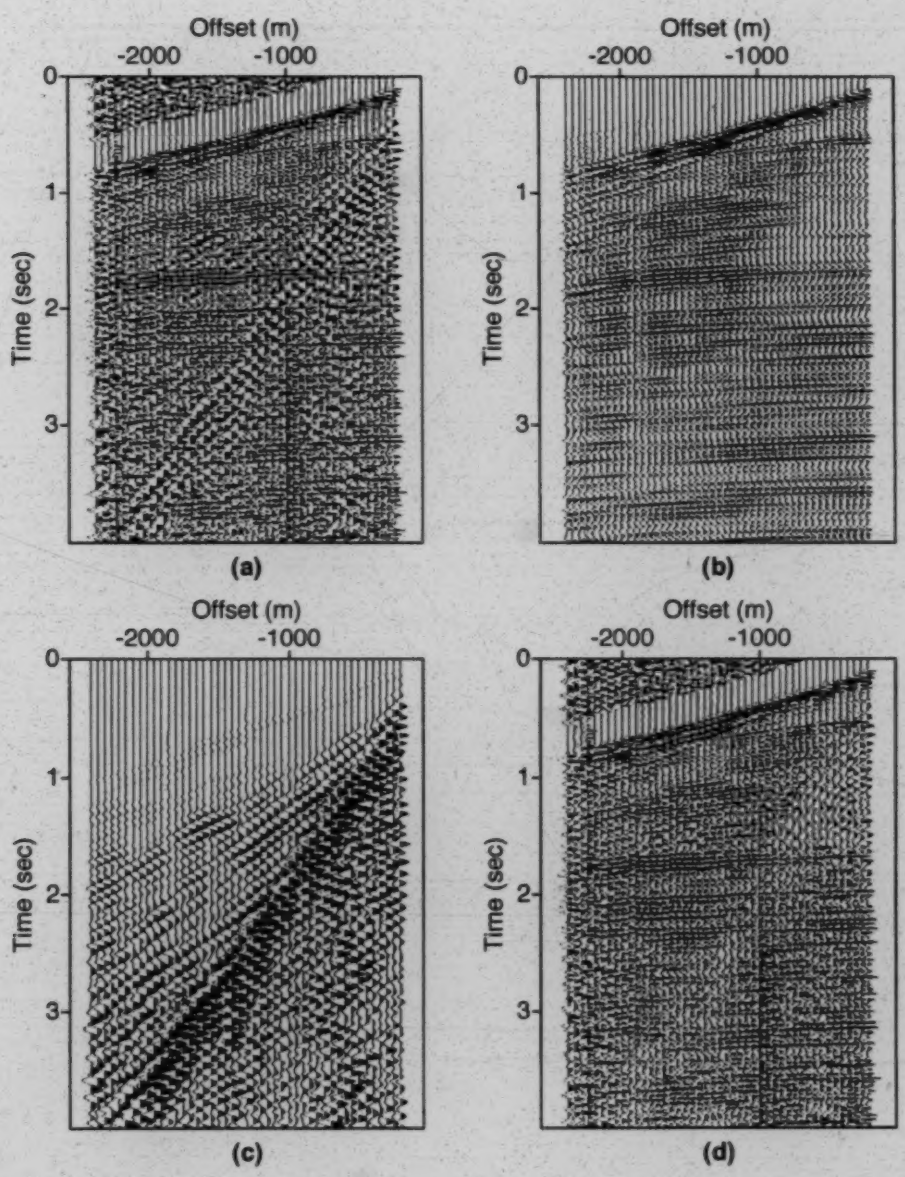

Figure 5.9: (a) Field data record containing reflections, ground roll, direct arrival, and random noise. (b) Recovered hyperbolic events from the hyperbolic DLS panel in Figure 5.8(b). (c) Recovered linear events from the linear DLS panel in Figure 5.8(b). (d) Data without ground roll. The linear events predicted in (c) are subtracted form the original data (a).

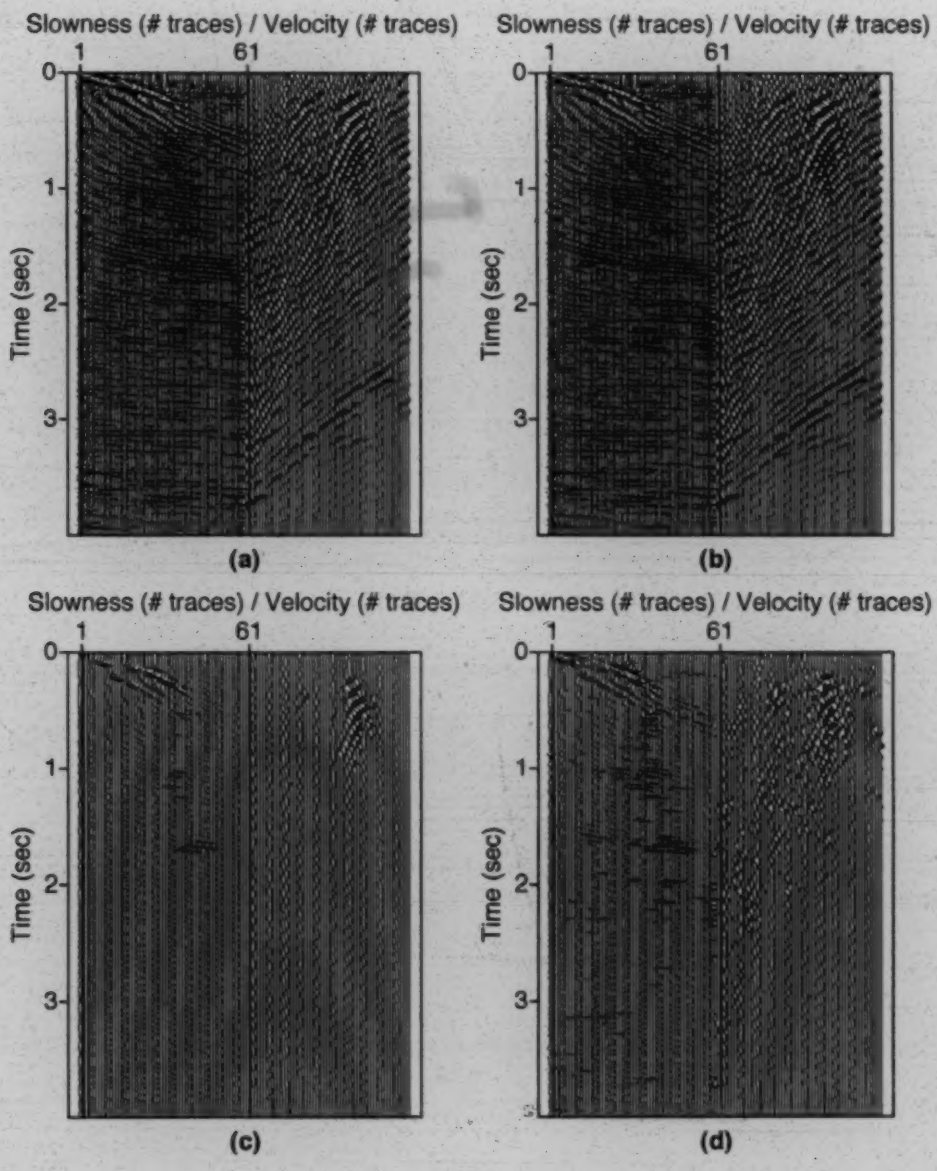

Figure 5.10: (a) DLS hybrid Radon panel after 5 iterations. (b) DLS hybrid Radon panel after 30 iterations. (c) IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. (d) IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$.

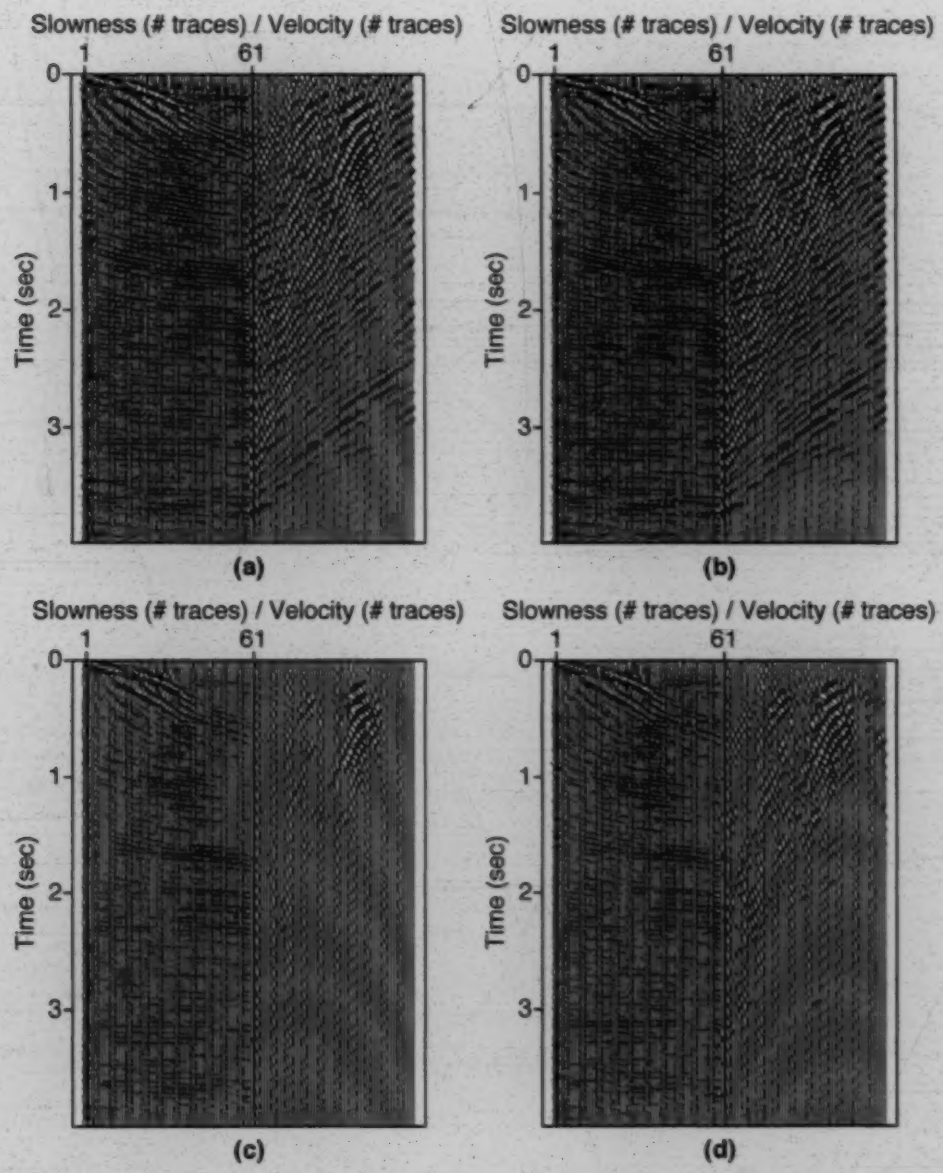

Figure 5.11: (a) DLS hybrid Radon panel after 5 iterations. (b) DLS hybrid Radon panel after 30 iterations. (c) IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. (d) IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$.

If we compare Figures 5.10(a)-(d) we can observe that the Radon panel in Figure 5.10(c), the IRLS solution with two external iterations and five internal iterations, is more sparse than the Radon panel in Figure 5.10(d) which represents the IRLS solution after two external iterations and thirty internal iterations. In addition, the IRLS solution in Figure 5.10(d) is more sparse than the DLS solutions in Figures 5.10(a) and (b). The conclusion drawn from this comparison is that the internal iterations play the role of the regularizer, assuring a stable solution that fits the data. The same can be observed in Figures 5.11(a)-(d).

The convergence of the IRLS algorithm, as mentioned in Chapter 4, can be achieved by using an optimum number of internal and external iterations. These tests are run to study the applicability of the algorithm in practical conditions. In this regard, we investigate the filtering using limited number of iterations and different trade-offs. Figures 5.10(c) and (d) illustrate the IRLS solutions obtained using a hyper-parameter $\sigma_m^2 = 0.1max|m_{DLS}|$ while Figures 5.11(c) and (d) illustrate the IRLS solutions obtained using a hyper-parameter $\sigma_m^2 = 0.5max|m_{DLS}|$. Comparing these figures, one can observe that when a small hyper-parameter is used a more sparse solution is obtained.

Figures 5.12 and 5.13 present the corresponding predicted data, Figures 5.14 and 5.15 represent the respective predicted linear data, while Figures 5.16 and 5.17 illustrate the data after the predicted linear events have been subtracted from the original data. By analyzing Figures 5.12(a)-(d) which represent represent the predicted data from the DLS and IRLS solutions from Figures 5.10(a)-(d) using a hyper-parameter $\sigma_m^2 = 0.1max|m_{DLS}|$. Because the IRLS solution obtained after two external iterations and five internal iterations (Fig. 5.10(c)) is too sparse, the predicted data in Figure 5.12(c) are also sparse, representing only a very small portion of the original data (Fig. 5.8(a)). On the other hand, the data predicted from the DLS solution reconstruct the original data almost entirely, including random noise. These conclusions are strengthen by the Figures 5.12-17.

Depending on the application, by analyzing images like Figures 5.8-5.17 one can make some decisions regarding the hyper-parameter σ_m^2, the number of internal iterations as well as the number of the external iterations. For this data set and the application of ground roll elimination I consider that $\sigma_m^2 = 0.1max|m_{DLS}|$, two external iterations and thirty internal iterations give an optimum result. I will now apply the same parameters

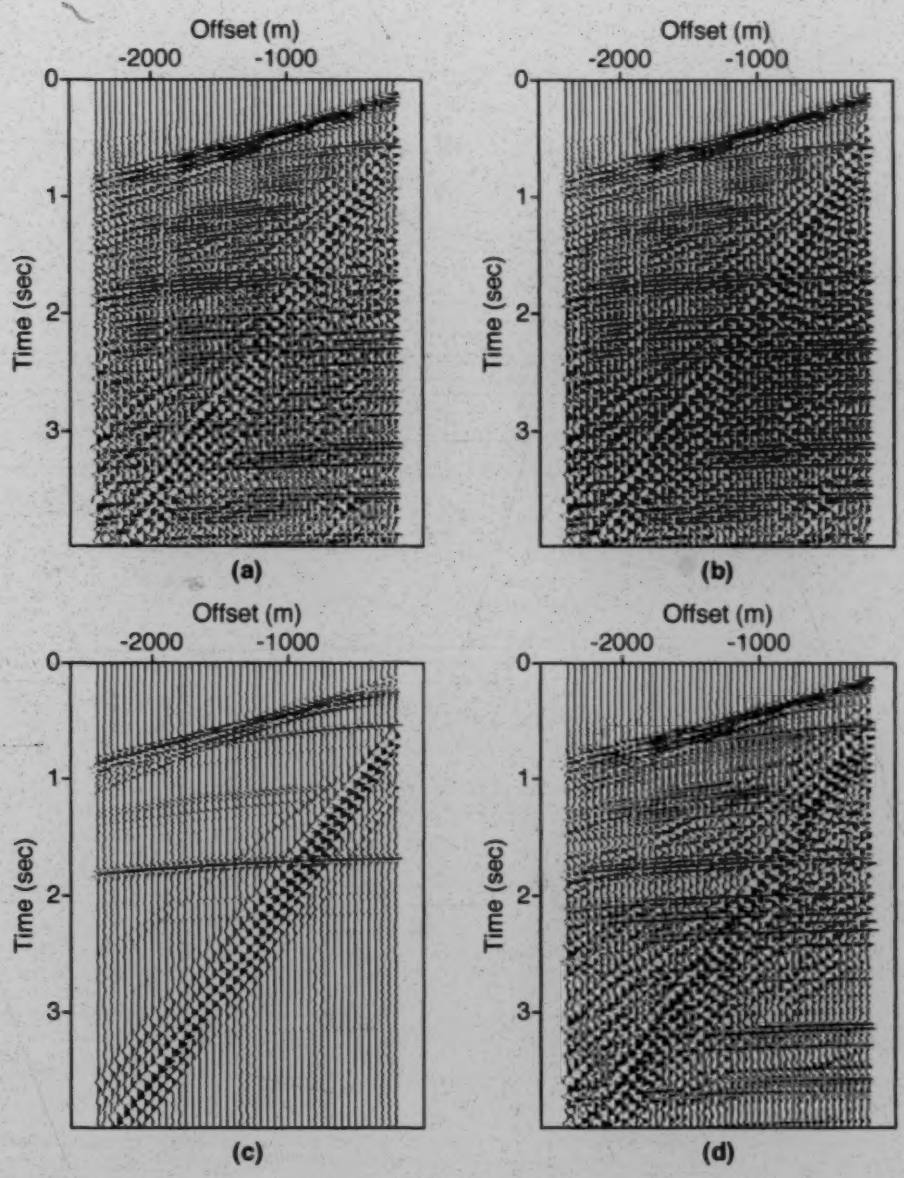

Figure 5.12: (a) Predicted data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. (d) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$.

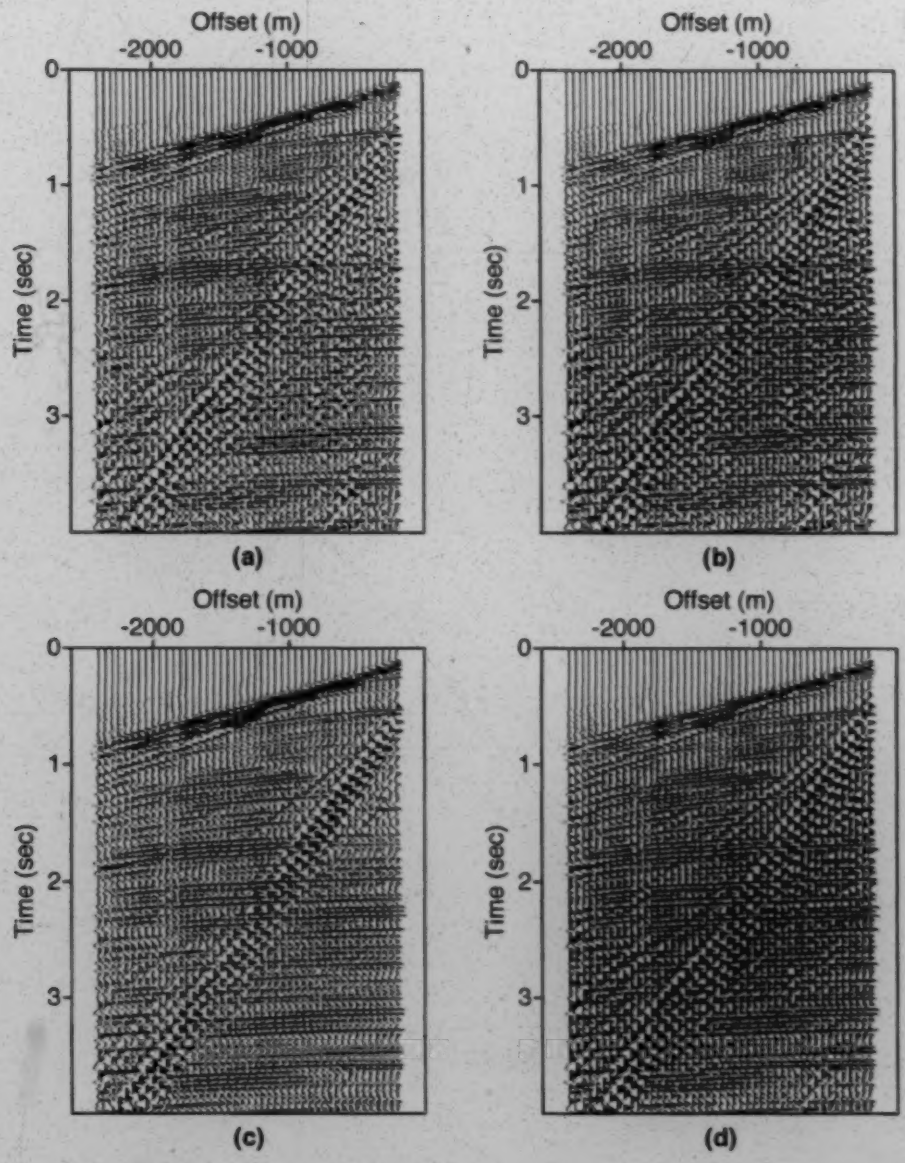

Figure 5.13: (a) Predicted data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. (d) Predicted data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$.

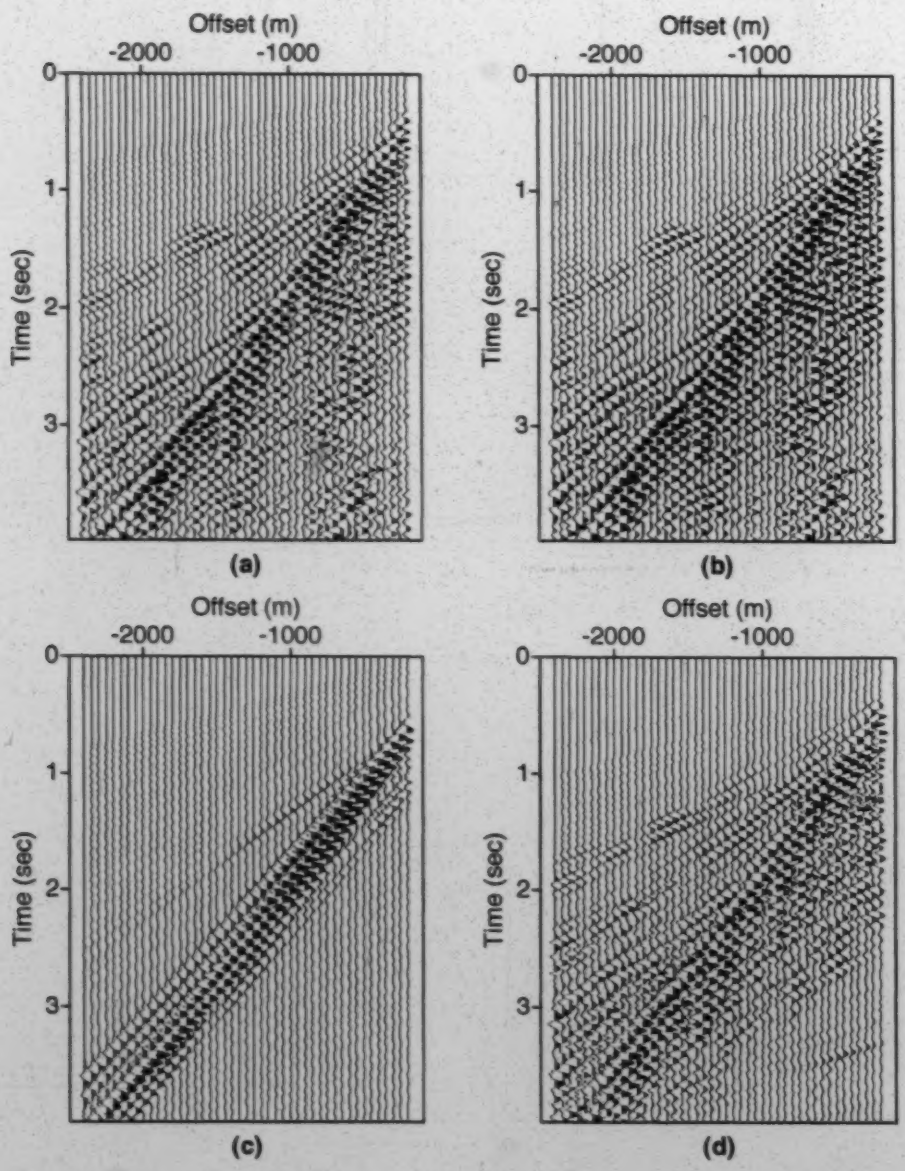

Figure 5.14: (a) Predicted linear data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted linear data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$. (d) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.1max|m_{DLS}|$.

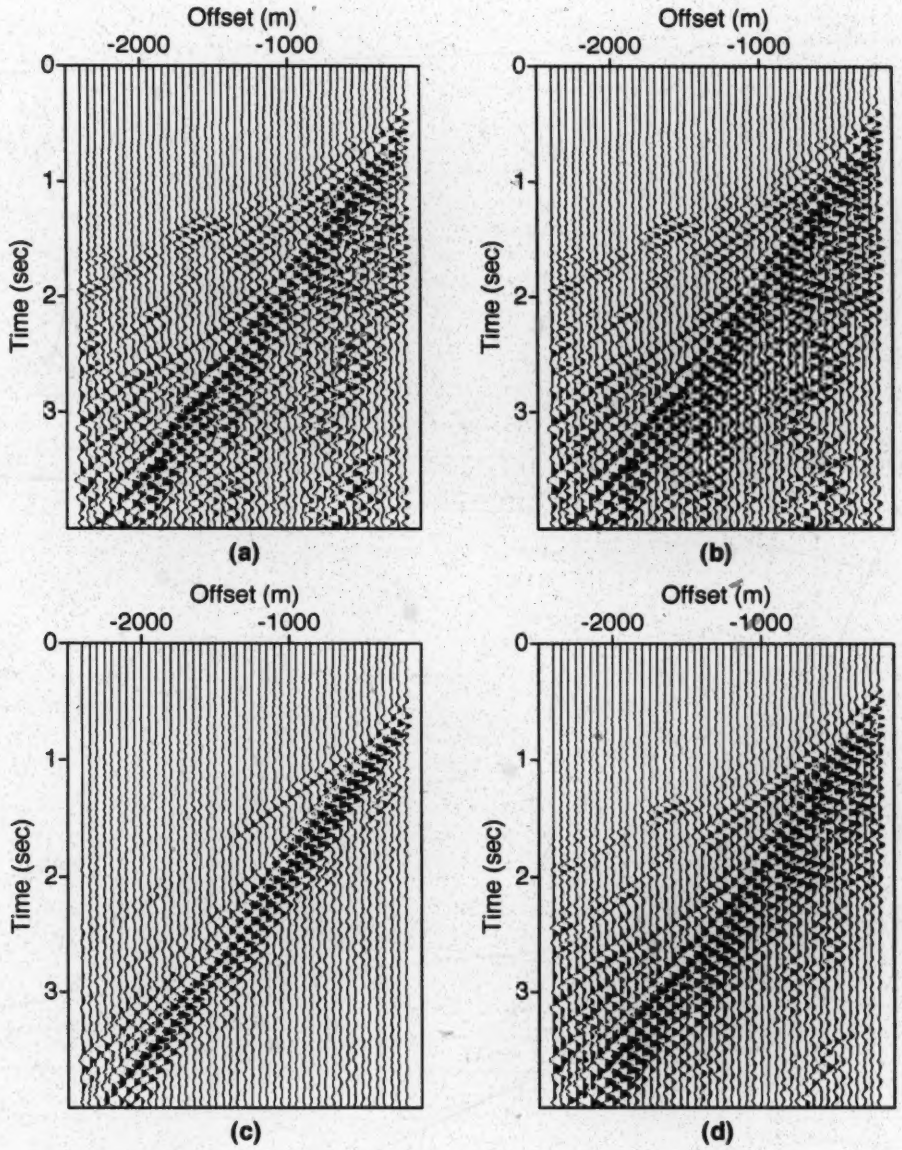

Figure 5.15: (a) Predicted linear data from the DLS hybrid Radon panel after 5 iterations. (b) Predicted linear data from the DLS hybrid Radon panel after 30 iterations. (c) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 5 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$. (d) Predicted linear data from the IRLS hybrid Radon panel after 2 external iterations and 30 internal iterations. $\sigma_m^2 = 0.5max|m_{DLS}|$.

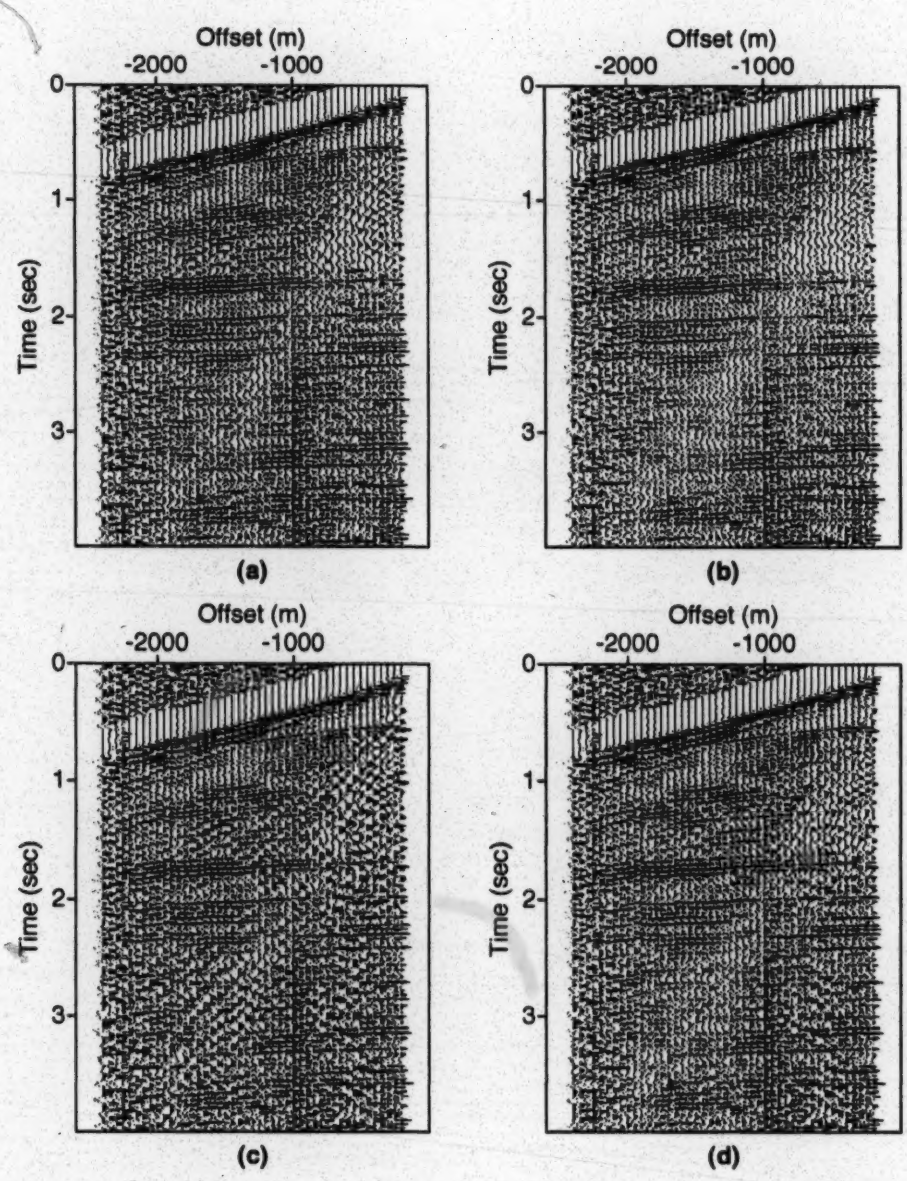

Figure 5.16: Ground roll attenuated data obtained by subtracting: (a) the predicted linear data (Fig. 5.14(a)) from the original data (Fig. 5.8(a)); (b) the predicted linear data (Fig. 5.14(b)) from the original data (Fig. 5.8(a)); (c) the predicted linear data (Fig. 5.14(c)) from the original data (Fig. 5.8(a)); (d) the predicted linear data (Fig. 5.14(d)) from the original data (Fig. 5.8(a)). $\sigma_m^2 = 0.1max|m_{DLS}|$.

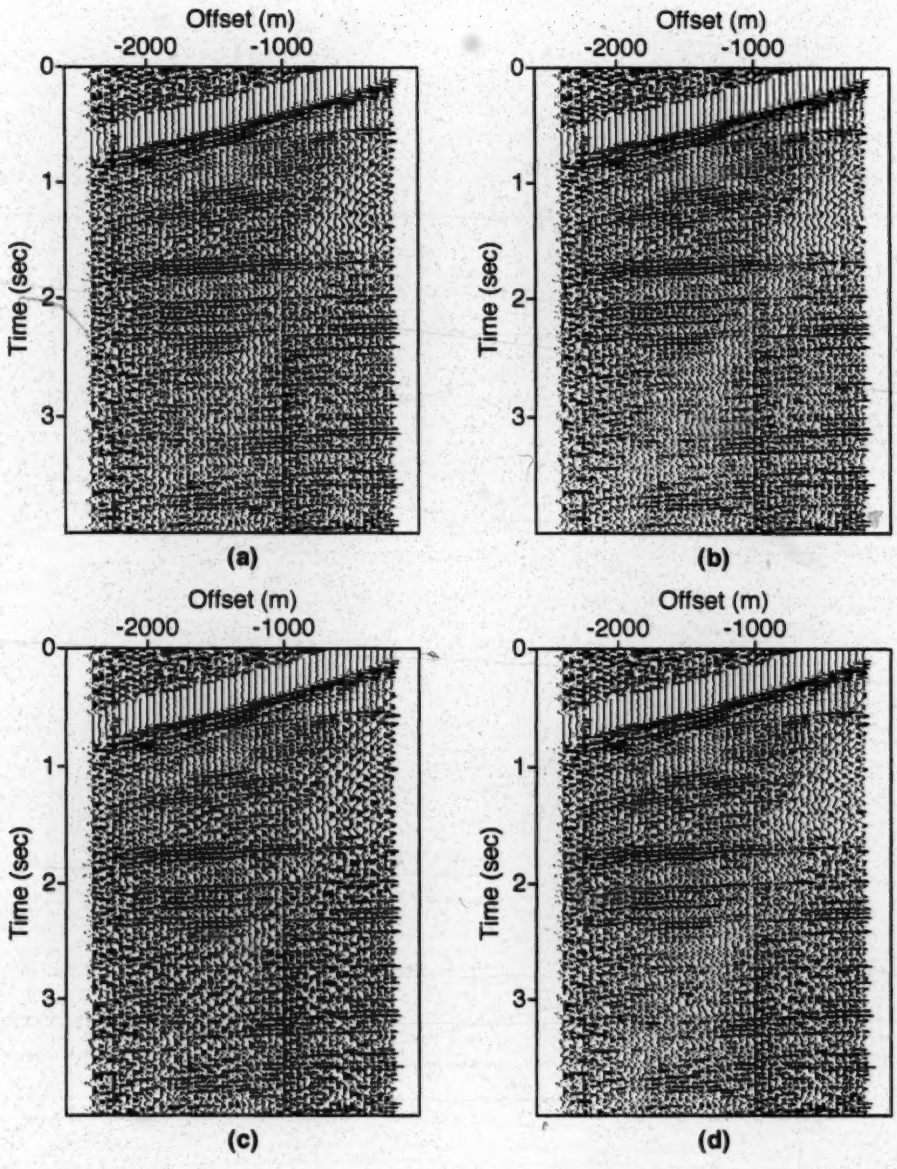

Figure 5.17: Ground roll attenuated data obtained by subtracting: (a) the predicted linear data (Fig. 5.15(a)) from the original data (Fig. 5.8(a)); (b) the predicted linear data (Fig. 5.15(b)) from the original data (Fig. 5.8(a)); (c) the predicted linear data (Fig. 5.15(c)) from the original data (Fig. 5.8(a)); (d) the predicted linear data (Fig. 5.15(d)) from the original data (Fig. 5.8(a)). $\sigma_m^2 = 0.5max|m_{DLS}|$.

to another seismic record and compare the result with the result obtain using an $f-k$ filter.

The second real data example shows a land record from North Africa represented by the shot number 10 from Yilmaz (1987) (Fig. 5.18(a). The source is a vibroseis, source that typically generates surface waves as ground roll. This record contains two strong shallow reflectors at about 0.8 sec and 1.1 sec obscured by the ground roll energy. There is also linear side-scattered energy at about 2.5 sec due to inhomogeneities in the subsurface. Figure 5.18(b) presents the IRLS hybrid Radon panel after thirty CG iterations and two up-dates. Because of the side-scattered energy that shows up as propagating at higher velocities than the ground roll and to avoid aliasing along the Radon parameter axis I increased the number of samples in Radon domain. Choosing the optimum range and sampling rate in Radon domain is a decisive factor in the quality of the result. Figures 5.18(c) and 5.18(d) represent the corresponding predicted linear events and the ground roll attenuated record obtained by subtracting the predicted linear events from the input data. It can be noticed that the linear energy was significantly attenuated.

Frequency-wavenumber filtering is the most common method applied in the seismic industry for coherent noise attenuation. Figure 5.19 illustrates the application and the result of a frequency-wavenumber filter to the same shot gather from North Africa. The data is firstly transformed to the $f-k$ domain (Fig. 5.19(b)) where a dip filter was applied. Figure 5.19(c) illustrates the filtered data in $f-k$ domain and time-offset domain (Fig. 5.19(d)). Figure 5.20(a) illustrates the difference between the original data and the filtered data, in other words the filtered noise. The $f-k$ spectrum of the noise is shown in Figure 5.20(b). Comparing the results of the hybrid Radon transform (Fig. 5.18(d)) and the $f-k$ filtering (Fig. 5.19(d)) we can notice that both methods give good results, managing to attenuate the linear events from the original data. The subject of this thesis is not to propose a new method for ground roll elimination, but rather to test a linear-hyperbolic hybrid Radon transform implemented in time domain as well as to investigate the influence of the number of external and internal iterations, Cauchy regularization, and trade-off parameter on the results.

To conclude, I applied a hybrid linear-hyperbolic Radon transform for ground roll attenuation to a third real data shot gather, this time from Middle East. The record is

record 39 from Yilmaz (1987), the source is dynamite (Geoflex) and it contains shallow reflections, ground roll, and short-period multiples (Fig. 5.21(a)). An IRLS hybrid Radon transform with thirty CG iterations and two up-dates is applied and the result is shown in Figure 5.21(b). The linear predicted data and the ground roll attenuated data are presented in Figures 5.21(c) and 5.21(d). The hyper-parameter used is again ten percent from the DLS solution with maximum amplitude.

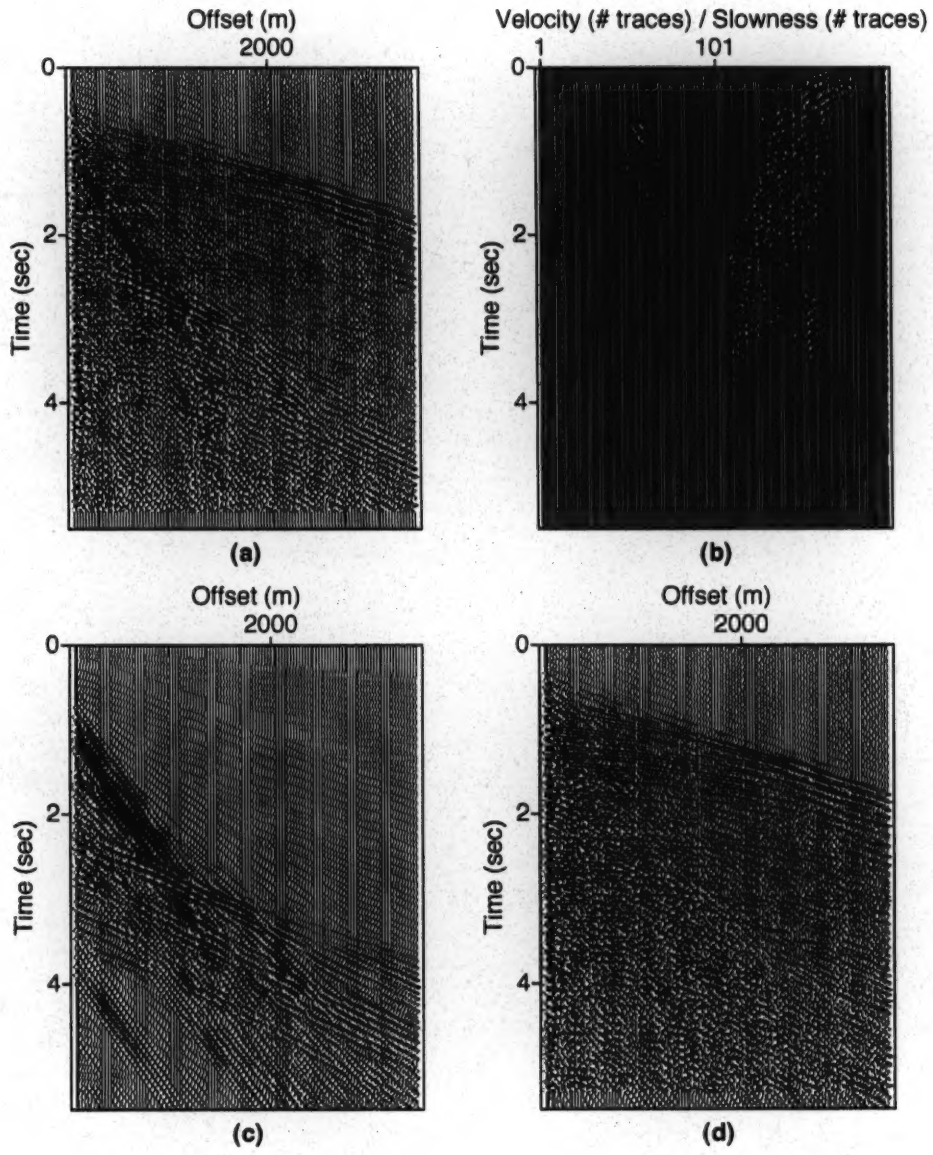

Figure 5.18: (a) Land data from North Africa, record number 10 from Yilmaz (1997). (b) IRLS hybrid Radon panel after 2 external iterations and 30 CG iterations every time. $\sigma_m^2 = 0.1max|m_{DLS}|$. (c) Predicted linear data from the linear panel from (b). (d) Ground roll attenuated data obtained from subtracting the linear predicted data (c) from the original data (a).

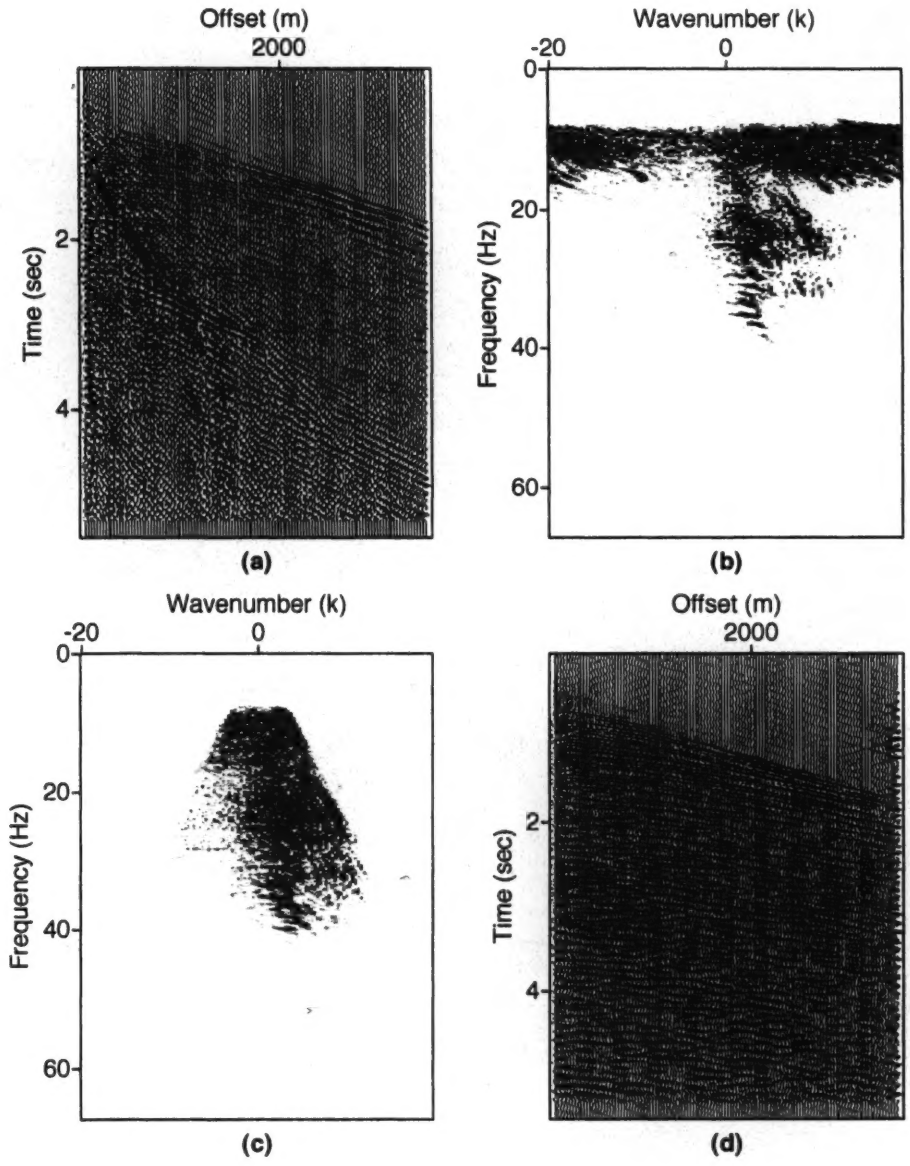

Figure 5.19: $f - k$ filtering of land data. (a) Land data before filtering. (b) $f - k$ spectrum of the data before filtering. (c) $f - k$ spectrum after the filter was applied. (d) Filtered data.

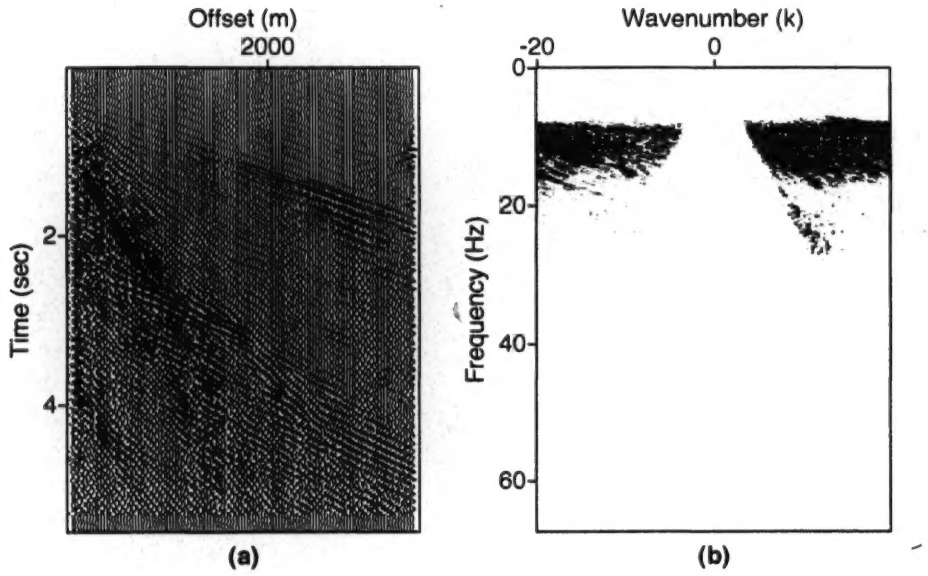

Figure 5.20: $f - k$ filtering of land data. (a) The filtered noise (Fig. 5.19(a)-Fig. 5.19(d)). (b) $f - k$ spectrum of the filtered noise.

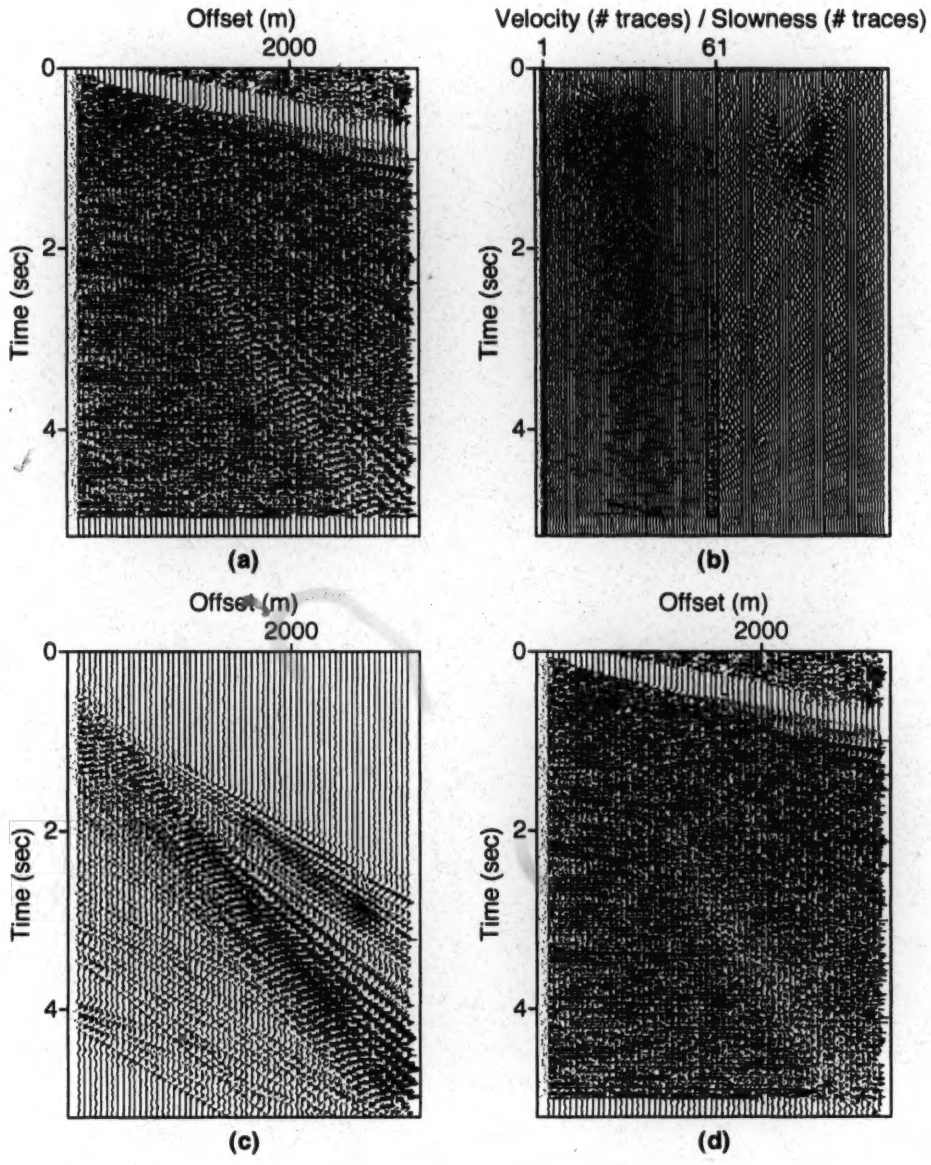

Figure 5.21: (a) Land data from Middle East, record number 39 from Yilmaz (1997). (b) IRLS hybrid Radon panel after 2 external iterations and 30 CG iterations every time. $\sigma_m^2 = 0.1max|m_{DLS}|$.(c) Predicted linear data from the linear panel from (b). (d) Ground roll attenuated data obtained from subtracting the linear predicted data (c) from the original data (a).

5.5 Summary

The separation of signal from coherent noise (ground roll) is performed in time domain using a hybrid linear-hyperbolic Radon transform. It is shown that by mapping separate events to different model spaces using a combined operator we can then model either one of the events by using only its forward operator. However, it is still preferable to model the linear noise and then subtract it from the input data, in order to preserve features of the signal that could be used in further data processing (i.e. amplitudes for azimuth versus offset (AVO) analysis). Introducing *a priori* information about the model is also improving the quality of the data in Radon domain. Sparseness constraints are introduced via Cauchy regularization norms and they are implemented via an iterative reweighted least squares (IRLS) algorithm explained in Chapter 4. Synthetic data examples were used to compare the damped least squares (DLS) and the high resolution solutions, and also to illustrate the performance of the hybrid Radon transform when applied to data contaminated with random noise. A real data example was used to illustrate the significance of the hyper-parameter σ_m^2. A second field data example is used to compare two ground roll attenuation methods, hybrid radon transform and frequency-wavenumber filtering. The third real data example is used to show one more time the result of the hybrid Radon transform when applied for ground roll suppression.

Chapter 6

Shifted Hyperbolic Radon Transform

6.1 Introduction

Currently, there is an increase in use of 2-D and 3-D wide-offset surveys in exploration seismology (Colombo, 2005). Long-offset seismic data provide significant illumination for deep reflections, especially in areas with complex structural tectonic regimes, such as Canadian foothills, or areas characterized by sharp lateral velocity changes (Dell'Aversana et al., 2003). While it improves the imaging of deeper targets like sub-basalt or sub-salt targets, some of the approximations commonly used in conventional processing, such as hyperbolic moveout, do not hold anymore. The Radon techniques have been efficiently applied for velocity estimation and analysis, as well as for multiple suppression, ground roll removal, and data interpolation. Resolution in Radon domain depends on mainly two factors: the basis function for the summation path, and the inversion algorithm. The use of the appropriate basis function for the summation curve yields a more focused image in Radon domain making the separation of events an easier task.In this chapter I present the problem of incorporating far offset approximation into the design of Radon transformations for velocity analysis. In particular, I explore the incorporation of the shifted hyperbola formula in the current implementation of time variant Radon transforms, and develop a framework for a multiparameter Radon transform. Resolution in Radon domain can be further improved by imposing a regularization constraint in the inversion algorithm (i.e. by introducing sparseness in the solution via a non-quadratic Cauchy norm). I use a synthetic data example to test the viability of a multiparameter Radon transform.

6.2 Methodology

Dix (1955) introduced the normal moveout (NMO) equation for a horizontally layered-earth model

$$t = \sqrt{t_0^2 + \frac{h^2}{V_{rms}^2}} \tag{6.1}$$

where t is the traveltime from the source to the receiver, t_0 represents the two-way vertical traveltime from the surface to the reflector, h is the offset, and V_{rms} notes the root-mean-square velocity. This is a short offset approximation and represents the first two terms of a Taylor's series expansion of t^2 around $h = 0$. For long offset we need to take into account at least one extra term of the expansion. Taner and Koehler (1969) give the following equation for the traveltime for a horizontally layered-earth model

$$t^2 = c_1 + c_2 h^2 + c_3 h^4 + c_4 h^6 + ... \tag{6.2}$$

where

$$c_1 = t_0^2 \tag{6.3}$$

$$c_2 = \frac{1}{\mu_2} \tag{6.4}$$

$$c_3 = \frac{1}{4}\frac{\mu_2^2 - \mu_4}{t_0^2 \mu_2^4} \tag{6.5}$$

$$c_4 = \frac{2\mu_4^2 - \mu_2\mu_6 - \mu_2^2\mu_4}{t_0^4 \mu_2^7} \tag{6.6}$$

Equation (6.2) represents an exact Taylor's series expansion of t^2 as a function of offset h. The coefficients μ_i are given by the following equation

$$\mu_i = \frac{\sum_{k=1}^{N} \Delta\tau_k V_k^i}{\sum_{k=1}^{N} \Delta\tau_k} \tag{6.7}$$

where $\Delta\tau_k$ is the vertical traveltime in the kth layer, and V_k is the interval velocity of the kth layer. It can be seen that $\mu_2 = V_{rms}^2$.

Two important characteristics of any equation used in seismic processing is accuracy and practicality. Although equation (6.2) is an accurate description of the traveltime at long offset it may not be practical to use it in data processing. Malovichko (1978, 1979) derived the shifted hyperbola NMO equation. Castle (1994) describes this equation for a horizontally layered earth as

$$t = t_0(1 - \frac{1}{S}) + \sqrt{\frac{t_0^2}{S^2} + \frac{h^2}{SV_{rms}^2}} \tag{6.8}$$

where

$$S = \frac{\mu_4}{\mu_2^2} \tag{6.9}$$

is a dimensionless parameter called the shift parameter. From the definition of S and μ_i it can be seen that for the first layer $S = 1$ equation (6.8) reduces to the normal moveout equation (6.1) (Dix, 1955). Making use of the Jensen inequality (Claerbout, 1992) it can be proved that $S \geq 1$.

$$\sum_{j=1}^{N} w_j f(p_j) - f(\sum_{j=1}^{N} w_j p_j) \geq 0. \tag{6.10}$$

If one considers $w_j = \Delta\tau_j$ and $f(p_j) = p_j^2 = v_j^4$ then the above inequality reduces to $S \geq 1$.

The Radon transform (RT) maps events with different curvatures in data domain, like primaries and multiples, to points in Radon domain. Due to this property, the Radon transform has been effectively used in coherent noise filtering, and data interpolation.

The forward Radon transform is defined as the following summation

$$d(h,t) = \sum_{v} m(v, \tau = \phi(v,t,h)) \tag{6.11}$$

where $d(h,t)$ is data in time domain, $m(v,\tau)$ represents the data in Radon domain, and $\phi(v,t,h)$ is the summation path - in this case, equation (8).

The summation can be written in matrix form as

$$\mathbf{d} = \mathbf{Lm} \tag{6.12}$$

where $\mathbf{L}$ is the Radon operator. By applying the adjoint operator $\mathbf{L^T}$ one can obtain the data in Radon domain as

$$\hat{\mathbf{m}} = \mathbf{L}^T\mathbf{d}. \tag{6.13}$$

It has been noticed by several authors (Thorson and Claerbout, 1985; Hampson, 1986; Kostov, 1990; Sacchi and Ulrych, 1995) that the utilization of equation (6.13) leads to a low resolution Radon panels. A key aspect in trying to circumvent the aforementioned problem entails defining the Radon transform as the solution of an inverse problem that can be solved by means of a conjugate gradient algorithm. Giving the following objective function

$$J = ||\mathbf{Lm} - \mathbf{d}||_2^2 + \mu||\mathbf{m}||_2^2 \tag{6.14}$$

and minimizing it with respect to $\mathbf{m}$ we obtain the least squares (LS) solution

$$\hat{\mathbf{m}} = (\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^T\mathbf{d}. \tag{6.15}$$

Most common integration paths used for Radon transform are linear, hyperbolic and parabolic but different basis functions can be adopted (i.e. elliptical Radon transform (Trad et al., 2002). In this chapter I present the advantages of using the shifted hyperbola equation as an integration path for velocity estimation from long-offset seismic data.

6.3 Synthetic data examples

Figure 6.1 shows a two layer earth model used to illustrate the results. The model is similar to the one used in Castle (1994). Offset is ranging from zero to 10 km with a sampling interval of 50 m. The depth of the second reflector is at 2 km, giving a maximum offset-to-depth ratio of 5.

The synthetic data gather calculated from the model is shown in Figure 6.2. Two summation curves for the second reflection are superimposed upon the data. While both the hyperbolic and shifted hyperbolic curves are a good approximation for offset-to-depth ratio up to one, the shifted hyperbola gives better results at far offsets (Fig. 6.2) given the fact that the right shift parameter is used.

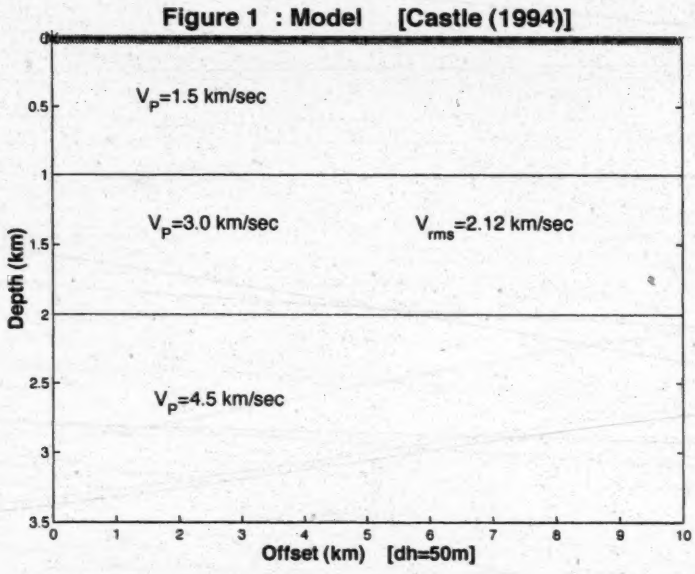

Figure 6.1: Three horizontal layered model (Castle, 1994). The parameters used for modeling are $V_{P1} = 1.5$ km/sec, $V_{P2} = 3$ km/sec, and $V_{P3} = 4.5$ km/sec.

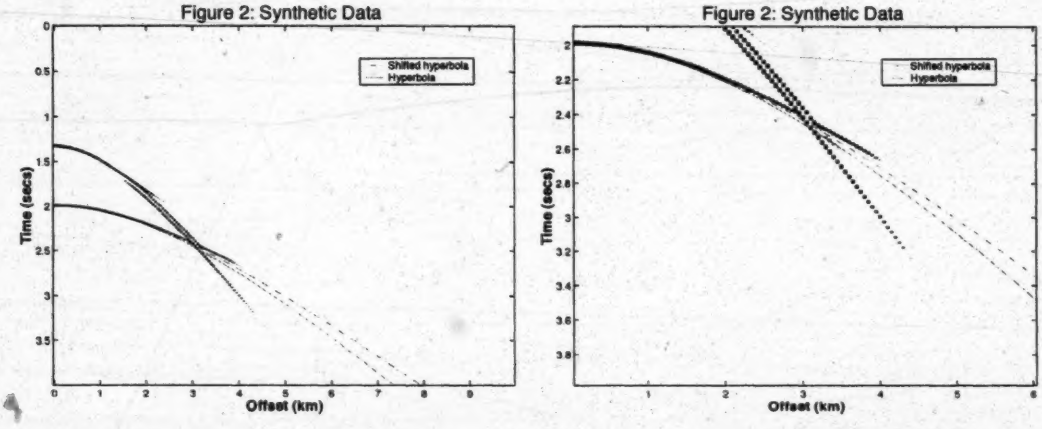

Figure 6.2: The synthetic data calculated from the model presented in the previous figure. Right side shows a zoom on the second reflection to emphasize the difference between the shifted hyperbola curve and hyperbola curve.

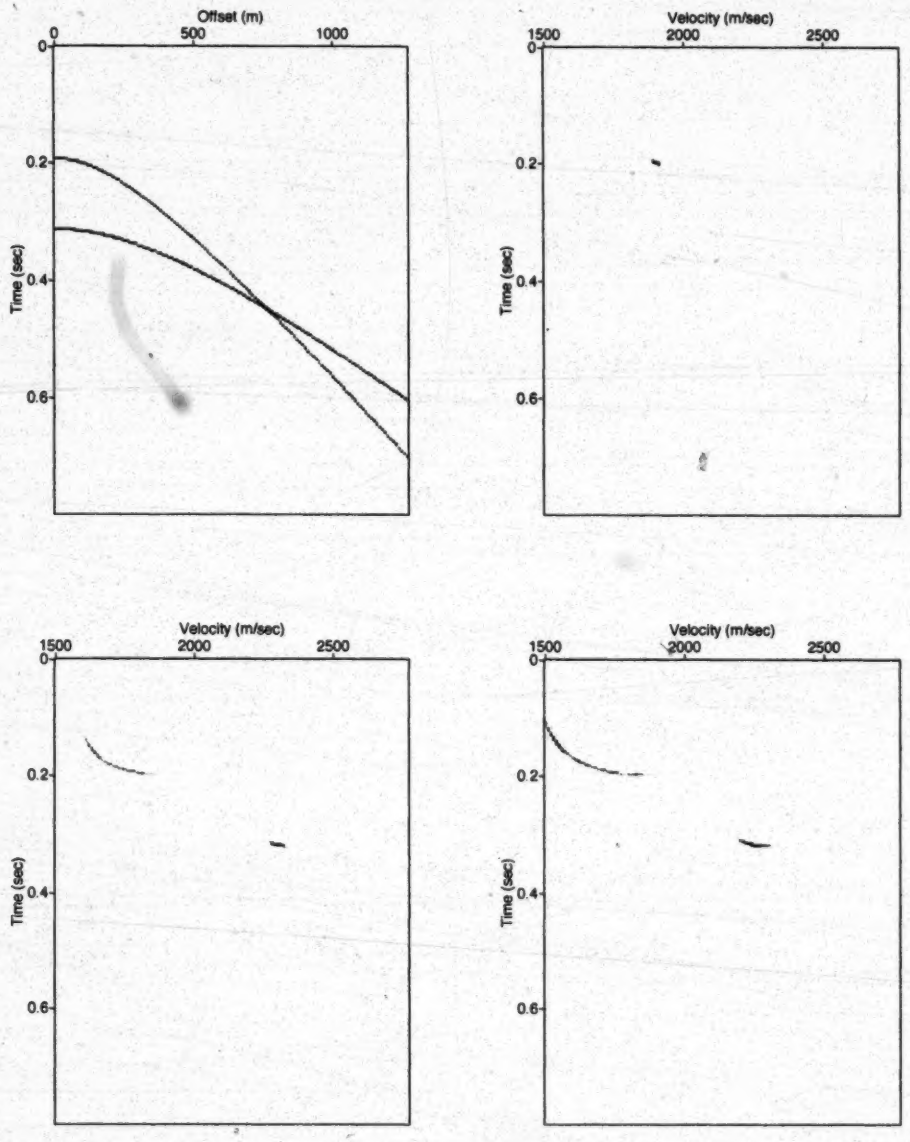

Figure 6.3: Synthetic data from a two layered model are shown in the upper left figure. Upper right figure shows the Radon panel obtained using HRT. The lower left figure shows the Radon panel obtained using shifted HRT (S=1.5) and the lower right figure shows the Radon panel obtained using S=1.7.

The second example illustrates the importance of the shift parameter (Fig. 6.3). The data were modeled using the shifted hyperbola equation. The two reflections are calculated from a model with two horizontal layers lying over a half-space. The two layers have velocities of 1900 m/sec and 2300 m/sec, respectively, and the shift parameter used is $S = 1$ for the first event and $S = 1.5$ for the second event. When the accurate shift parameter is used, the image in Radon domain is more focused and the correct velocity is obtained (Fig. 6.3(b) and 6.3(c)). Significant smearing of reflections occurs when the wrong parameter is considered and the velocity is different than the true one. When S is smaller than the true value, the obtained velocity is larger than the true velocity and frown-shaped smearing occurs (Fig. 6.3(b)). When S is larger than the true shift parameter, the obtained velocity will be too low and smile-shaped smearing occurs (Fig. 6.3(c) and 6.3(d)). In this example it has been observed that the shift parameter S is sensitive to changes of 0.1 in its value.

Figure 6.4 shows a synthetic data shot gather that has been modeled by the shifted hyperbola equation using the parameters in Table 6.1. The offset is ranging from 0 to 5 km with a sampling interval of 0.02 km. The depth of the last reflector is 1.7 km, giving a maximum offset-to-depth ratio of approximately 2.

Figure 6.5 illustrates several shifted hyperbolic Radon panels with constant shift parameter (S) obtained using a different S for each of them.

A shifted hyperbolic Radon transform using only a constant S gives an approximate estimation of the shift parameter for different reflectors. However, the focusing in Radon domain is achieved only for the events for which S is close to the right one. Figure 6.6 illustrates the principle of a time variant shifted hyperbolic Radon transform. A more global approach includes the use of a time dependent shift parameter. This transform would allow to scan for velocity while tuning the shift parameter. Figure 6.7 shows several models obtained by applying a shifted hyperbolic Radon transform with the shift parameter linearly varying with time

$$S(\tau) = S_0 + a\tau. \tag{6.16}$$

A priori information about S, such as the range of variation, is helpful in choosing the intercept S_0 and the slope a. As shown in de Vries and Berkhout (1984) and Sacchi et al.

Thickness (km)	Interval velocity (km/s)	S
0.150	1.500	1.00
0.100	1.800	1.03
0.200	2.200	1.11
0.150	2.350	1.12
0.125	2.500	1.13
0.075	2.800	1.15
0.043	3.200	1.18
0.050	3.500	1.24
0.040	3.750	1.29
0.020	3.850	1.31
0.080	4.000	1.40
0.100	4.300	1.50
0.110	4.500	1.57
0.150	4.800	1.64
0.172	5.000	1.67
0.100	5.200	1.69

Table 6.1: Modeling parameters for the synthetic data presented in Figure 6.4.

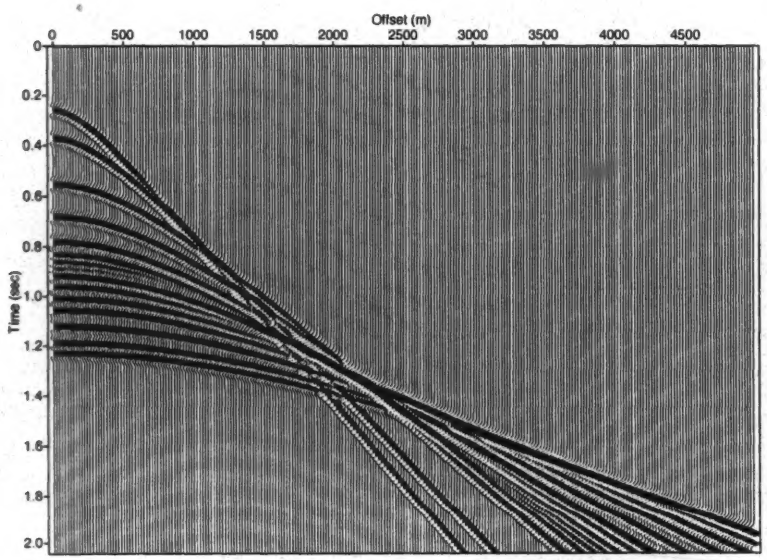

Figure 6.4: The synthetic data calculated from the model presented in Table 6.1 .

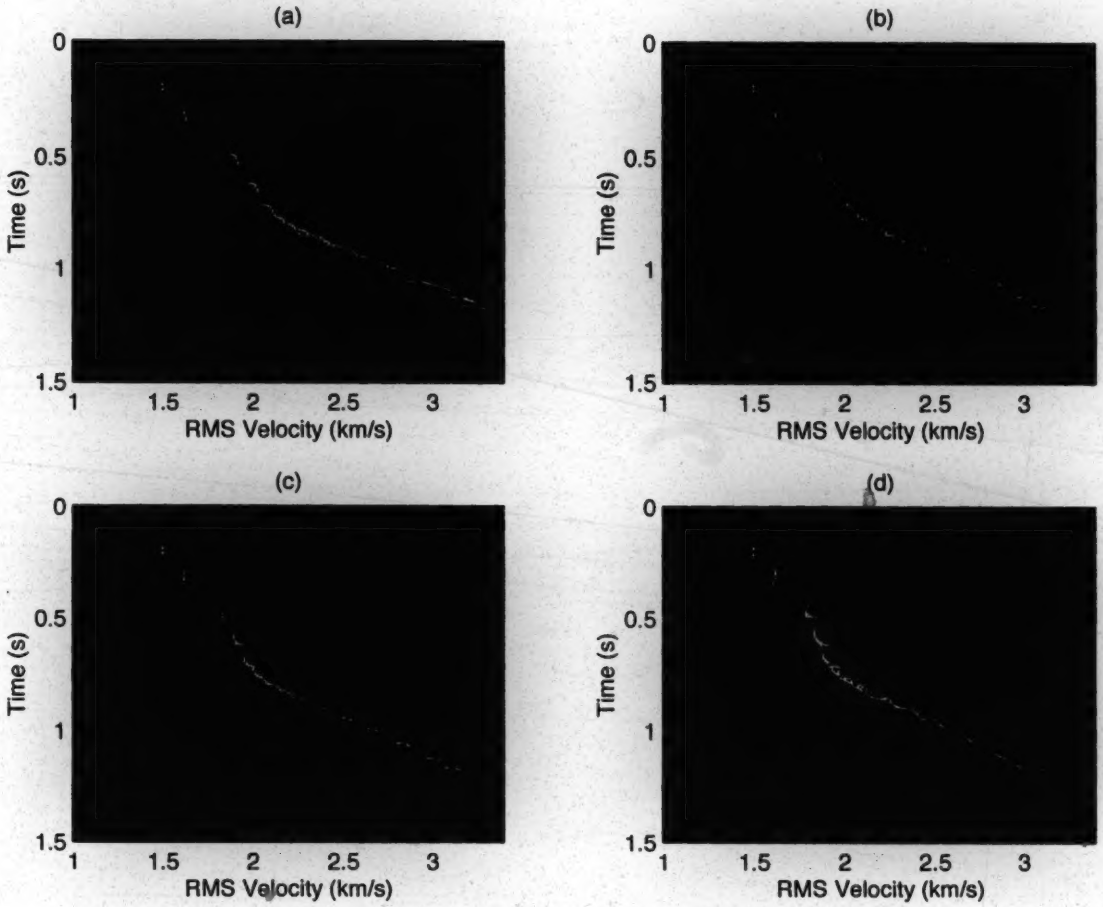

Figure 6.5: Shifted hyperbolic Radon panels with constant shift parameter (S). (a) Hyperbolic Radon panel (equivalent to shifted hyperbolic Radon with constant $S = 1$). (b) Shifted hyperbolic Radon panel for $S = 1.3$. (c) Shifted hyperbolic Radon panel for $S = 1.4$. (d) Shifted hyperbolic Radon panel for $S = 1.7$.

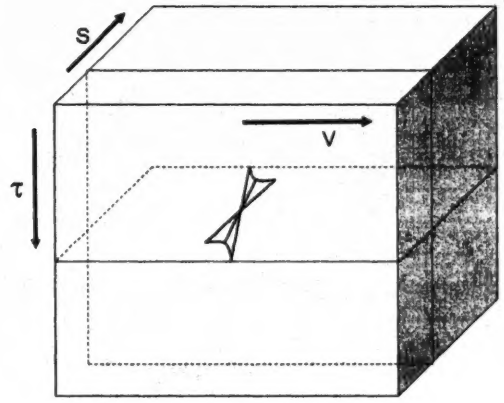

Figure 6.6: Diagram that illustrates the principle of a time-variant shifted hyperbolic Radon transform.

(1996), minimum entropy norms can be used as a measure of resolving power. Therefore, the most focused model is obtained by minimizing the entropy which is equivalent to maximizing the following function called negentropy

$$E = \frac{1}{N \ln N} \sum_{i=1}^{M} q_i \ln q_i \tag{6.17}$$

where

$$q_i = N \frac{m_i^2}{\sum_{k=1}^{N} m_k^2}, \tag{6.18}$$

N is the size of the model, and m_i represents Radon panel for a particular $S(\tau)$ curve. The maximum value of negentropy is equivalent to the maximum focusing in Radon domain, as shown in Figure 6.8 and 6.7(b). Figure 6.9 illustrates the maximum negentropy principle. In the first case, Figure 6.9(a), it is shown a sparse model, with all samples zero except one whose amplitude is N (the number of samples). When computing the negentropy E, one will obtain $E = 1$. As the degree of sparseness in the model decreases, the

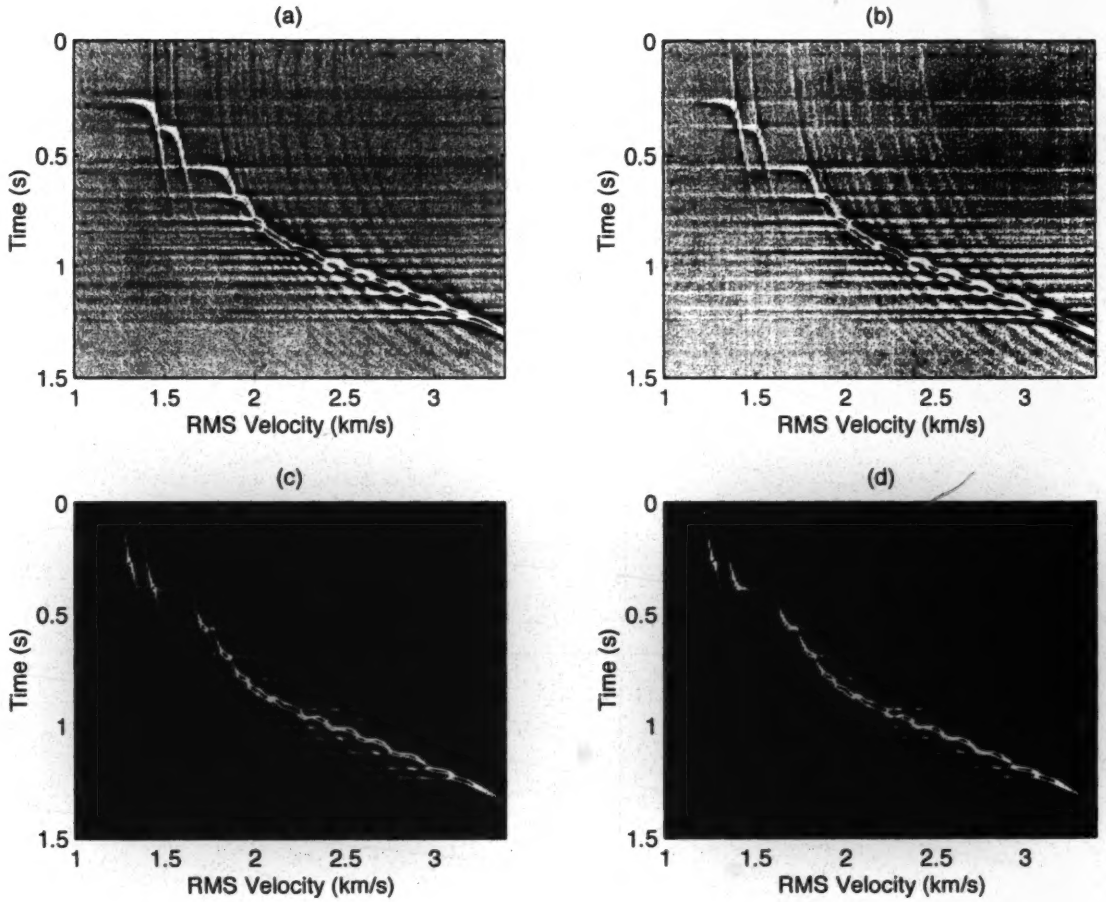

Figure 6.7: Shifted hyperbolic Radon panels with variable shift parameter (S). Model 1 ($S(t) = 1$). (b) Model 5 (most focused $a = 0.002$). (c) Model 6. (d) Model 8.

negentropy decreases as well. The other extreme situation is the last case, Figure 6.9(d), in which all the samples have the same amplitude equal to 1/N, giving a negentropy $E = 0$.

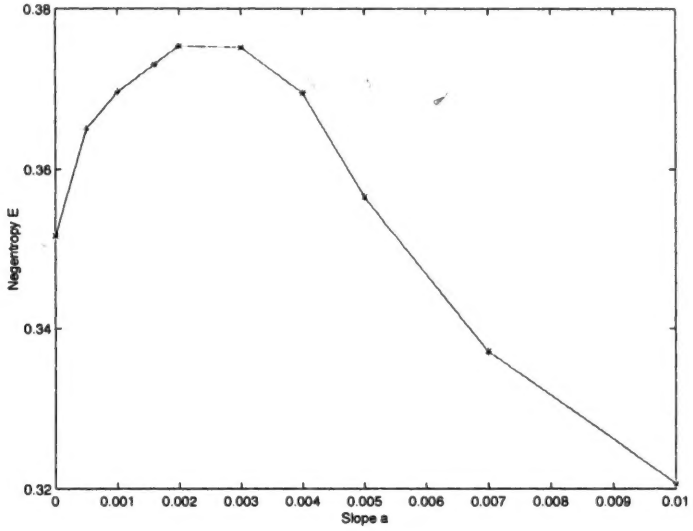

Figure 6.8: Focusing measure curve for the 10 models using a different slope.

6.4 Summary

For long-offset data the shifted hyperbola represents a more accurate approximation. An extra unknown parameter called shift parameter is introduced. In this chapter I modify the current implementation of the hyperbolic Radon transform to incorporate the shifted hyperbola formula (Castle, 1994). The quality of the results strongly depends on the shift parameter. A correct value of the parameter yields a more focused image in the Radon domain and the obtained velocity is close to the true velocity. We firstly estimate a range for the shift parameter S by applying the shifted hyperbolic Radon transform with constant S. Subsequently, a shifted hyperbolic Radon transform with variable $S(\tau)$ is applied. The most focused image in Radon domain is chosen as being the model with the maximum value of negentropy. In this synthetic example, simple linear $S(\tau)$ curves give good results, but more complicated curves can be later incorporated. To further increase

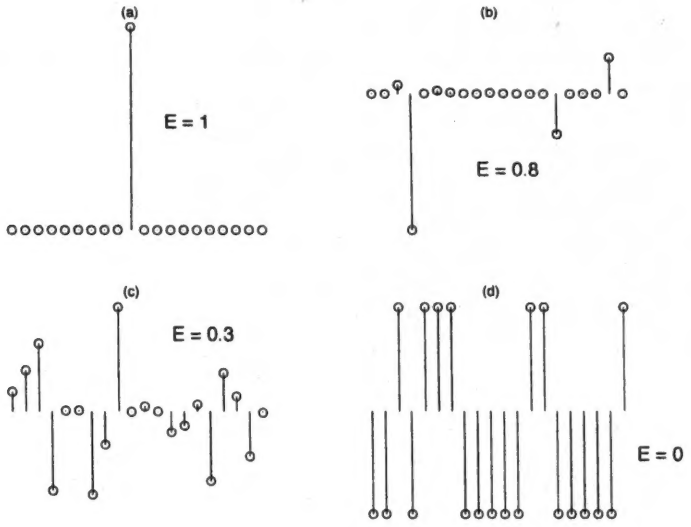

Figure 6.9: Focusing measure for the shifted hyperbolic Radon transform.

the focusing in the Radon domain, model-dependent non-quadratic regularization via a Cauchy norm can be introduced.

Chapter 7

Discussion and conclusions

Seismic methods are the most popular geophysical methods used in hydrocarbon exploration. Noise, random or not, represents a constitutive and at the same time unwanted part of the seismic data. Random noise is typically the result of problems in the equipment, weather conditions or environment and it can often be eliminated in the acquisition step. Coherent noise is generally the result of the subsurface conditions and it is more difficult to remove in the acquisition step. Developing techniques to eliminate ground roll and multiple reflections from seismic data has been a ceaseless subject of interest in exploration seismology.

In this thesis, I have presented Radon techniques to remove coherent noise from seismic data. Ground roll and multiple reflections have been defined in Chapter 2 and typical methods to eliminate them have been reviewed. I have briefly described frequency and frequency-wavenumber filtering methods and applied them for ground roll attenuation. Signal and noise bandwidth may overlap in which case frequency based attenuation methods offer suboptimal results. Spatial and temporal aliasing due to poor sampling in time and offset are a few of the disadvantages of $f - k$ filtering. Other methods have been also mentioned for ground roll elimination. Multiples suppression techniques can be classified into filtering and prediction methods and have been equally described.

Radon transform, commonly used in CAT scanning, image processing, remote sensing, and geophysics, has been defined in Chapter 3. It can be classified in time-invariant (i.e linear, and parabolic Radon transform) and time-variant (i.e. hyperbolic Radon transform). Time-invariant transforms are often implemented in frequency domain to increase

the computational speed. Time-variant transforms can be implemented only in the time domain but allow more flexibility and accuracy in choosing the basis functions. Both implementations are explained in this thesis. However, because all the applications involved time-variant transforms, I used only the time domain version.

I posed Radon transform as an inverse problem and solved it by an iterative reweighted least squares conjugate gradient algorithm (IRLS). Solving Radon transform as an inverse problem allows one to incorporate *a priori* information about the solution. For noise attenuation problem the goal is to obtain focused projections of the data in Radon domain. This can be achieved by imposing sparseness constraints in our solution. I included a Cauchy regularization norm in the cost function to take into account the sparseness of the model. I adopted a right preconditioning technique to increase the computational efficiency of the conjugate gradient least squares (CGLS) algorithm. All these concepts are detailed in Chapter 4.

In Chapter 5 I described hybrid Radon transform (Trad et al., 2001) as a method to eliminate ground roll. In this thesis, the transform is a combination of linear and hyperbolic Radon transform, therefore implemented in time domain as opposed to the combination of linear and pseudo-hyperbolic Radon transform encountered in Trad et al. (2001) and implemented in frequency domain. The advantage of the time domain implementation is the use of hyperbolas as basis function which is a much better approximation of the reflections. Linear and hyperbolic events contained in a common shot or common midpoint (CMP) gather will be transformed to linear and hyperbolic Radon domain respectively. The computational time is the same with applying only one of the transforms and no filtering in Radon domain is required. Either ground roll or reflections can be recovered by applying the corresponding forward transform. Synthetic and real data examples are used to illustrate the performance of hybrid Radon for ground roll attenuation. One important condition to apply hybrid Radon successfully is that the basis functions should be significantly different so that events map to different domains. The hyperbolic-linear hybrid Radon works well for near to intermediate offset data. For long offset data, the hyperbola asymptotically resembles lines, therefore part of the energy belonging to the reflections will map to the linear Radon domain. This phenomenon is typically known as crosstalk and could be alleviated by introducing additional constraints

in the inverse problem.

In Chapter 6 I adopted the shifted hyperbola equation (Castle, 1997) as a basis function for the Radon transform to improve the approximation of the reflection events at long offset. The shifted hyperbolic Radon transform is implemented in the time domain and applied on synthetic data for velocity analysis and estimation. A future application of this transform is considered to be suppression of multiple reflections.

7.1 Future work

A problem that still remains unsolved in hybrid Radon transform is crosstalk, the phenomenon of contaminating the linear Radon panel with energy belonging to the hyperbolic Radon panel. A possible approach consists in adopting a Bayesian framework for inversion. The weighting functions are linked to the covariance matrix which is a measure of correlation (Menke, 1984), in this case correlation between the models. Uncorrelated models give rise to diagonal covariance matrix and the problem reduces to the problem presented in this thesis. Correlated models, give rise to non-diagonal covariance matrix, and the off-diagonal elements are closely connected to the degree of overlapping of the model domains. The problem consists in developing an algorithm to solve the new system of equations and implementing it. If this problem is overcome, the hybrid Radon transform can be generalized to include more than two basis functions.

Velocity analysis and estimation using the shifted hyperbolic Radon transform should be tested with real data. Application to multiple suppression in long offset seismic data should be also considered.

References

Al-Yahya, K., 1991, Application of the partial Karhunen-Loeve transform to suppress random noise in seismic sections: Geophys. Prosp., **39**, 77-93.

Alliney, S., and Ruzinsky, S. A., 1994, An algorithm for the minimization of mixed l_1 and l_2 norms with application to Bayesian estimation: IEEE Trans. Signsl Processing, **42**, 618-627.

Alleon, G., Benzi, M., and Giraud, L., 1997, Sparse approximate inverse preconditioning for dense linear systems arising in computational electromagnetics: Numerical Algorithms, **16**, 1-15.

Anstey, N. A., 1986, Whatever happened to ground roll?: Leading Edge, **5**, no 3, 40-45.

Benzi, M., 2002, Preconditioning Techniques for Large Linear Systems: A Survey: Journal of Computational Physics, **182**, 418-477.

Berkhout, A. J. and Verschuur, D. J., 1997, Estimation of multiple scattering by iterative inversion, part I: Theoretical considerations: Geophysics, **62**, 1586-1595.

Berryhill, J. R. and Kim, Y. C., 1986, Deep-water peg-legs and multiples - Emulation and suppression: Geophysics, Soc. of Expl. Geophys., **51**, 2177-2184.

Beylkin, G., 1987, Discrete Radon Transform: IEEE transactions on acoustics, speech, and signal processing, **35**, 162-172.

Brysk, H., Cathriner, M., Goodrum, R. A. and Pennacchioni, J. L., 1987, Predictive deconvolution of cylindrical slant stacks, 57th Ann. Internat. Mtg: Soc. of Expl. Geophys., Session:S17.4.

Calvetti, D., 2006, Preconditioned iterative methods for linear discrete ill-posed problems from a Bayesian inversion perspective: Journal of Computational and Applied Mathematics, *In Press*, available on line.

Cary, P., 1998, The simplest discrete Radon transform: Soc. of Expl. Geophys., 68th Ann. Internat, Mtg., 1999-2002.

Castle, R. J., 1994, A theory of normal moveout: Geophysics, **59**, 983-999.

Challenor, P. G., Cipollini, P., and Cromwell, D., 2001, Use of the 3-D Radon Transform to examine the properties of oceanic Rossby waves: Journal of Atmospheric and Oceanic Technology, **18**, no. 9, 1558-1566, 2001. (PDF file - 1256k). See also: 2002, Corrigendum: JAOT, **19**, no. 5, 828.

Chapman, C. H., 1978, A new method for computing synthetic seismograms: Geophys. J. R. Astr. Soc., **54**, 481-518.

Chen, S. S., Donoho, D. L., and Saunders, M. A., 1998, Atomic decomposition by basis pursuit: SIAM Journal on Scientific Computing, **20**, 1, 33-61.

Claerbout, J. F. , 1992, Earth sounding analysis, Processing versus inversion: Blackwell Scientific Publications, Inc.

Colombo, D., 2005, Benefits of Wide Offset seismic for commercial exploration targets and implications for data analysis: The Leading Edge, 352-363.

Deighan, A.J., and Watts, D.R., 1997, Ground roll suppression using the wavelet transform: Geophysics, **62**, 1896-1903.

Dell'Aversana, P., Colombo, D., Buia, M., and Morandi, S., 2003, Velocity/interface model building in thrust belt by tomographic inversion of global offset seismic data, Geophysical Prospecting, **51**, 23-35.

Dix, C. H. , 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68-86.

Foster, D. J., and Mosher, C. C., 1992, Suppression of multiple reflections using the Radon transform: Geophysics, **57**, 3, 386-395.

Gomez Londono, E., Castillo Lopez, L., and de Souza Kazmierczak, 2005, Using the Karhunen-Loeve transform to suppress ground roll in seismic data: Earth Sci. Res. J., **9**, 2, 139-147.

Greenhalgh, S. A., Mason, I. M., Lucas, E., Pant, D., and Eames, R. T., 1990, Controlled direction reception filtering of P- and S-waves in tau-p space: Geophys. J. Int, **100**, 221-234.

Gulunay, N., 1990, F-X domain least-squares tau-p and tau-q, 60th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1607-1610.

Haber, E., 1997, Numerical strategies for the solution of inverse problems: University of British Columbia, PhD Thesis, Vancouver, BC.

Hampson, D. , 1986, Inverse velocity stacking for multiple elimination: Journal of the Canadian Society of Exploration Geophysicists, **22**, 44-55.

Hampson, D., 1987, The discrete Radon transform: a new tool for image enhancement and noise suppression: 57th Ann. Internat. Mtg: Soc. of Expl. Geophys., Expanded Abstracts, 141-143.

Hanke, M., 1995, Conjugate Gradient Type Methods for Ill-Posed Problems: Longman.

Hanke, M., and Hansen, P., 1993, Regularization methods for large scale problems: Survey. Math. Ind., **3**, 253-315.

Hansen, P. C., 1998, Rank-Deficient an Discrete Ill-Posed Problems: Numerical aspects of Linear Inversion. SIAM monographs on mathematical modeling and computation.

Hestenes, M., and Stiefel, E., 1952, Method of conjugate gradients for solving linear systems: J. Res. Natl. Bur. Stand., **49**, 409–436.

Hindriks, C. O. H., and Duijnadam, A. J. W., 1998, Radon domain reconstruction of 3-D irregularly sampled VSP data: Soc. of Expl. Geophys., 68th Ann. Internat, Mtg., 2003-2006.

Huber, P. J., 1981, Robust Statistics: John Wiley and Sons, Inc.

Jones, I. F., and Levy, S., 1987, Signal-to-noise ration enhancement in multichannel seismic data via the Karhunen-Loeve transform: Geophys. Prosp., **35**, 12-32.

Kabir, M. M. N., and Verschuur, D. J., 1995, Groundroll removal using the linear Radon transform, 57th Mtg.: Eur. Assn. of Expl. Geophys., Session: A050.

Kak, A. C., and Slaney, M., 1988, Principles of Computerized Tomographic Imaging: IEEE Press.

Kommedal, J. H. and Tjostheim, B. A., 1989, Tutorial: A study of different methods of wavefield separation for application to VSP data: Geophys. Prosp., Eur. Assn. Geosci. Eng., 37, 117-142.

Kneib, G. and Bardan, V., 1997, 3D targeted multiple attenuation: Geophys. Prosp., Eur. Assn. Geosci. Eng., **45**, 701-714.

Kostov, C., 1990, Toeplitz structure in slant-stack inversion: 60th Ann. Internat. Mtg: Soc. of Expl. Geophys., Expanded Abstracts, 1618-1621.

Levin, S., 2003, Fast, effective curved ray moveout corrections for Kirchhoff time migration: 73rd Ann. Internat. Mtg: Soc. of Expl. Geophys., Expanded Abstracts.

Li, Q., 2001, High resolution hyperbolic Radon transform multiple removal: University of Alberta, M.Sc. Thesis, Edmonton, AB.

Lines, L. R. and Treitel, S., 1984, Tutorial - A review of least-squares inversion and its application to geophysical problems : Geophys. Prosp., Eur. Assn. Geosci. Eng., 32, 159-186. (* Discussion in GRP-38-1-101-103 with reply by authors).

Liu, X., 1999, Ground roll supression using the Karhunen-Loeve transform: Geophysics, **64**, 564-566.

Lokshtanov, D. E., 1993, Adaptive multiple suppression in tau-p domain: 63rd Ann. Internat. Mtg: Soc. of Expl. Geophys., 1086-1089.

Malovichko, A. A., 1978, A new representation of the traveltime curve of reflected waves in horizontally layered media: Applied Geophysics (in Russian), **91**, no. 1, 47-53.

Maroof, S. I. and Gravely, C. J., 1984, Aliasing and tau-p, 54th Ann. Internat. Mtg: Soc. of Expl. Geophys., Session:S14.8.

Menke, W., 1984, Geophysical data analysis: discrete inverse theory: Academic Press.

Moldoveanu-Constantinescu, C., and Sacchi, M. D., 2005, Enhanced resolution in Radon

domain using the shifted hyperbola equation: 75th Ann. Internat. Mtg: Soc. of Expl. Geophys., Expanded Abstracts.

Moon, W., Carswell, A., Tang, R., and Dilliston, C., 1986, Radon transform wavefield separation for vertical seismic profiling: Geophysics, **4**, 940-947.

Morse, P. F., and Hildebrandt, G. F., 1988, Ground roll suppression by the stackarray, 58th Ann. Internat. Mtg.: Soc. of Expl. Geophys., Session: S14.5.

Nemeth, T., Sun, H., and Schuster, G. T. , 2000, Separation of signal and coherent noise by migration filtering: Geophysics, **65**, 2, 574-583.

Nemeth, T., and Bube, K. P. , 2001, An operator decomposition approach for the separation of signal and coherent noise in seismic wavefields: Inverse Problems, **17**, 533-551.

Peacock, K. L., and Treitel, S., 1969, Predictive deconvolution - Theory and Practice, Geophysics, **34**, 155-169.

Radon, J., 1917, Uber die Bestimmung von Funktionen durch ihre Integralwerte langsgewisser Manningfaltigkeiten (On the determination of functions from their integrals along certain manifolds): Berichte Saechsische Akademie der Wissenschaften, **29**, 262-277.

Robinson, E. A., 1954, Predictive decomposition of time series with applications to seismic exploration: M.I.T. PhD Thesis, Cambridge, Mass.

Robinson, E. A., and Treitel, S., 1980, Geophysical signal analysis: Prentice-Hall, Inc.

Ryu, J. V., 1982, Decomposition (DECOM) approach applied to wave-field analysis with seismic reflection records: geophysics, **47**, 869-883.

Saad, Y., 1996, Iterative methods for sparse linear systems: PWS publishing Company.

Sacchi, M. D., 1996, Aperture Compensated Radon and Fourier Transforms: University of British Columbia, PhD Thesis, Vancouver, BC.

Sacchi, M. D., 1997, Reweighting strategies in seismic deconvolution: Geophysical Journal International, **129**, 651-656.

Sacchi, M. D., 2002, Karhunen-Loeve (KL) Filtering of Seismic Data Algorithm: Signal Analysis and Imaging Group (SAIG), Department of Physics, University of Alberta, Edmonton, AB.

Sacchi, M. D., and Ulrych, T. , 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, 4, 1169-1177.

Sacchi, M. D., Velis, D. R. and Cominguez, A. H., 1996, Minimum entropy deconvolution with frequency-domain constraints, in Robinson, E. A. and Osman, O. M., Ed., Deconvolution 2: Soc. of Expl. Geophys., 278-285.

Sacchi, M. D., and Porsani, M. J., 1999, Fast high resolution Radon transform: 69th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1657-1660.

Scales, J. A., 1987, Theory of Seismic Imaging: Samizdat Press.

Scales, J. A., and Gersztenkorn, A., 1988, Robust methods in inverse theory: Inverse Problems, **4**, 1071-1091.

Scales, J. A., Gersztenkorn, A., and Treitel, S., 1988, Fast l_p solution of large, sparse, linear systems: Application to seismic travel time tomography: Journal of Computational Physics, **75**, 314-328.

Schultz, P. S.,and Claerbout, J. F., 1978, Velocity estimation and downward-continuation by wavefront synthesis: Geophysics, **43**, no. 4, 691-714.

Sheriff, R. E., 1991, Encyclopedic Dictionary of Exploration Geophysics, 3rd Edition: Soc. of Expl. Geophys., 384.

Stoffa, P. L., Diebold, J. B. and Buhl, P., 1982, Velocity analysis for wide-aperture seismic data: Geophys. Prosp., Eur. Assn. Geosci. Eng., 30, 25-57.

Strang, G., 1986, Introduction to applied mathematics: Wellesley-Cambridge Press.

Strutt, J. W. (Lord Rayleigh), 1885, On waves propagated along the plane surface of elastic solid: Proc. London Math. Society, **17**, 4-11.

Taner, M. T., 1980, Long-period sea-floor multiples and their suppression: Geophys. Prosp., Eur. Assn. Geosci. Eng., **28**, 30-48.

Taner, M. T., and Koehler, F., 1969, Velocity spectra-digital computer derivation and applications of velocity functions: Geophysics, **34**, 859-881.

Tarantola, A., 1987, Inverse Problem Theory: Methods for Data Fitting and Model Parameter estimation: Elsevier.

Tatham, R. H., Goolsbee, D. V., Massell, W. F., and Nelson, H. R., 1983, Seismic shear-wave observation in a physical model experiment: Geophysics, **48**, 688-701.

Tatham, R. H., and Goolsbee, D. V., 1984, Separation of P- and S-wave reflections offshore western Florida: Geophysics, **49**, 493-508.

Thorson, R., and Claerbout, J. , 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, **50**, 4, 2727-2741.

Trad, D., 2001, Implementations and applications of the sparse Radon transform: University of British Columbia, PhD Thesis, Vancouver, BC.

Trad, D. and Ulrych, T., 2000, A different approach to hyperbolic and elliptical Radon transforms, 70th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1949-1952.

Trad, D., Sacchi, M. D., and Ulrych, T. J., 2001, A hybrid linear-hyperbolic Radon transform: Journal of Seismic Exploration, **9**, 303-318.

Trad, D., Ulrych, T. J., and Sacchi, M. D., 2002, Accurate interpolation with high-resolution time-variant Radon transforms: Geophysics, **67**, 644-656.

Trad, D., 2002b, Interpolation with migration operators: 72nd Ann. Internat. Mtg: Soc. of Expl. Geophys.

Trad, D., Ulrych, T. J., and Sacchi, M. D., 2003, Latest views of the sparse Radon transform: Geophysics, **68**, 386-399.

Ulrych, T. J., Freire, S. L., and Siston, P., 1988, Eigenimage processing of seismic sections: 58th Ann. Internat. Mtg., Soc. Expl. Geophysics., Expanded Abstracts, 1261-1265.

Ulrych, T. J., Sacchi, M. D., and Woodbury, A., 2001, A Bayes tour of inversion: A tutorial: Geophysics, Soc. Expl. Geophysics., **66**, 55-69.

Verschuur, D. J. and Berkhout, A. J., 1997, Estimation of multiple scattering by iterative inversion, part II: Practical aspects and examples: Geophysics, **62**, 1596-1611.

de Vries, D. and Berkhout, A. J., 1984, Velocity analysis based on minimum entropy: Geophysics, **49**, 2132-2142.

Weglein, A. B., Carvalho, P. M. and Stolt, R. H., 1992, Nonlinear inverse scattering for multiple suppression: Application to real data. part I, 62nd Ann. Internat. Mtg: Soc. of Expl. Geophys., 1093-1095.

Wiggins, J. W., 1988, Attenuation of complex water-bottom multiples by wave equation-based prediction and subtraction: Geophysics, Soc. of Expl. Geophys., 53, 1527-1539.

Yilmaz, O., 1987, Seismic data processing: Soc. Expl. Geophys.

Yilmaz, O., 1989, Velocity stack processing: Geophys. Prosp., Eur. Assn. Geosci. Eng., **37**, 357-382.

Youzwishen, C. F., 2001, Non-linear sparse and blocky constraints for seismic inverse problems: University of Alberta, M.Sc. Thesis, Edmonton, AB.